Université Paris 1 Panthéon Sorbonne, École Doctorale de Philosophie

THÈSE

pour l'obtention du grade de docteur en philosophie

Présentée et soutenue publiquement par :

Marie Darrason

Le 2 juillet 2014

Y a-t-il une théorie génétique de la maladie ?

Sous la direction de Jean GAYON, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Composition du Jury :

Laurent ABEL	Directeur de recherche à l'INSERM
Giovanni BONIOLO	Professeur à l'Université de Milan, Italie
Jean GAYON	Professeur à l'Université Paris 1 Panthéon-Sorbonne
Maël LEMOINE	Maître de Conférences HDR à l'Université de Tours
Sandra MITCHELL	Professeur à l'Université de Pittsburgh, Pennsylvanie, USA
Michel MORANGE	Professeur à l'École Normale Supérieure de Paris

Université Paris 1 Panthéon Sorbonne, Ecole Doctorale de Philosophie

THÈSE

pour l'obtention du grade de docteur en philosophie

Présentée et soutenue publiquement par :

Marie Darrason

Le 2 juillet 2014

Y a-t-il une théorie génétique de la maladie ?

Sous la direction de Jean GAYON, Professeur à l'Université Paris 1 Panthéon-Sorbonne

Composition du Jury :

Laurent ABEL	Directeur de recherche à l'INSERM
Giovanni BONIOLO	Professeur à l'Université de Milan, Italie
Jean GAYON	Professeur à l'Université Paris 1 Panthéon-Sorbonne
Maël LEMOINE	Maître de Conférences HDR à l'Université de Tours
Sandra MITCHELL	Professeur à l'Université de Pittsburgh, Pennsylvanie, USA
Michel MORANGE	Professeur à l'École Normale Supérieure de Paris

À la mémoire de mes grands parents,

Jeanne Darrason et Daniel Berrut

Remerciements

On imagine souvent le doctorant en philosophie comme un être solitaire, isolé dans sa tour d'ivoire ou enfermé dans quelque bibliothèque. Rien ne saurait s'éloigner davantage de ce que j'ai vécu pendant mes années de doctorat. J'ai eu la chance de bénéficier d'un entourage exceptionnel à qui je veux exprimer ici toute ma gratitude.

Je remercie tout d'abord mon directeur de thèse, Jean Gayon. C'est en suivant son cours sur le concept de gène que j'ai pris goût à la philosophie de la biologie et que je me suis intéressée à la génétique médicale. Ses conseils chaleureux, ses encouragements constants et ses commentaires exigeants m'ont poussée à aborder avec toujours plus de rigueur et de précision les concepts scientifiques comme les concepts philosophiques. Je veux lui dire aussi toute ma reconnaissance pour la confiance et la liberté qu'il m'a accordées au cours de ces années de doctorat. C'était un pari que de donner sa chance et de soutenir la candidature d'une étudiante qui souhaitait aussi poursuivre ses études de médecine en parallèle de sa thèse. Mais Jean Gayon m'a aussi encouragée à profiter pleinement de toutes les expériences possibles qu'offre le doctorat, de la participation aux conférences au séjour à l'étranger. Il a enfin constamment soutenu et guidé, avec beaucoup de bienveillance et d'attention, les différents projets de philosophie de la médecine que nous lui avons soumis avec Hélène Richard, de l'ouverture du séminaire de philosophie de la médecine à l'organisation en juin 2013 de l'*International Advanced Seminar in Philosophy of Medicine*, - expériences qui m'ont permis de découvrir la communauté des philosophes de la médecine et qui se sont révélées déterminantes pour mon travail de thèse.

Je remercie chaleureusement Laurent Abel, Giovanni Boniolo, Maël Lemoine, Sandra Mitchell et Michel Morange pour avoir accepté de faire partie de mon jury de thèse. Je suis tout particulièrement reconnaissante à Sandra Mitchell et à Giovanni Boniolo d'avoir fait le voyage pour venir assister à ma soutenance.

L'IHPST a d'abord été le lieu où j'ai découvert la philosophie de la médecine et où je me suis formée à cette discipline jeune mais incroyablement dynamique, en

compagnie de tous ceux qui ont participé à cette belle aventure du Séminaire Philmed. J'ai pris un plaisir extraordinaire à nos échanges (parfois fiévreux !) et je remercie du fond du cœur les co-organisatrices, Hélène Richard et Cécilia Bognon-Küss, et tous les membres du séminaire, en particulier, Maël Lemoine, Jean-Baptiste Trabut, Smaïl Bouaziz, Arnaud Plagnol, Delphine Olivier, Franck Brouillard, Hidetaka Yakura, Hajimé Fujimori, Eric Charmetant et Isabelle Rémy-Largeau.

Je ne sais comment remercier Hélène Richard. Nous avons commencé notre doctorat et imaginé tous nos projets de philosophie de la médecine ensemble. Nous avons à ce point débattu de nos sujets que nos thèses se répondent : elle écrit la thèse d'éthique en génétique médicale que j'aurais voulu écrire. Je lui sais un gré infini de nos discussions, de nos cafés partagés et de son amitié.

Je tiens à exprimer toute ma reconnaissance à Maël Lemoine : j'éprouve un grand respect et beaucoup d'admiration pour sa façon de faire de la philosophie de la médecine, et son arrivée, en tant que chercheur associé à l'IHPST, a constitué un tournant dans mon travail de thèse. On reconnaîtra en particulier son influence dans la façon dont je considère que la philosophie de la médecine fait partie de la philosophie des sciences et dans la manière dont j'aborde le concept de théorie médicale. Nos discussions et ses encouragements m'ont été d'une aide précieuse.

Bien d'autres membres de l'IHPST ont contribué à ce travail. Je remercie Matteo Mossio pour avoir organisé le Club PhilBioMed. Je remercie Philippe Huneman pour ses conseils avisés. Je suis tout particulièrement reconnaissante à Thomas Pradeu, pour sa disponibilité, son écoute bienveillante et le défi qu'ont constitué ses objections. Je remercie Lucie Laplane, dont la thèse a été une source d'inspiration. Je remercie enfin Florent Franchette, Gauvain Leconte, Gladys Kostyrka et Gaëlle Pontarotti, avec qui j'ai eu beaucoup de plaisir à partager ces années de doctorat, ainsi qu'Armelle Thomas, Peggy Tessier et Sandrine Souraya, sans qui le laboratoire ne pourrait fonctionner.

Je remercie l'Ecole Doctorale de Paris 1, et tout particulièrement Chantal Jacquet, pour le soutien institutionnel et financier dont j'ai pu bénéficier, aussi bien d'un point de vue personnel dans le cadre de conférences ou de mon séjour à l'étranger, que du point de vue des projets que nous avons organisés en philosophie de la médecine avec Hélène Richard.

Je remercie Sandra Mitchell de m'avoir accueillie avec une incroyable générosité et une grande disponibilité au Département d'Histoire et de Philosophie des sciences de l'Université de Pittsburgh de septembre à décembre 2011. Nos rencontres hebdomadaires animées au sein du groupe « *Causes and Diseases* », en compagnie de Marta Bertolaso, Maria Kronfeldner et Lauren Ross, furent passionnantes et ses remarques, toujours chaleureuses, ont été déterminantes pour le plan de mon argumentation. Je suis également très reconnaissante à Ken Schaffner pour m'avoir permis de participer au groupe de lecture de son prochain livre et pour les dizaines d'articles de génétique qu'il m'a fait découvrir. Je remercie aussi Peter Machamer, Kathryn Tabb, Julia Bursten avec qui j'ai pris grand plaisir à interagir, ainsi que les organisateurs du WIP Seminar qui m'ont donné l'occasion de présenter et de discuter mes travaux.

Je remercie Laurent Abel pour avoir répondu avec une grande bienveillance à mes nombreuses questions sur la théorie génétique des maladies infectieuses.

J'adresse mes plus sincères remerciements à tous ceux qui m'ont fait l'amitié de relire un ou plusieurs chapitres de cette thèse et qui ont joué un rôle décisif dans la forme du manuscrit final, en particulier Elodie Giroux, Thomas Pradeu, Franck Brouillard, Hélène Richard, Smaïl Bouaziz et Delphine Olivier.

Mes amis ont été une source inépuisable d'énergie et de réconfort tout au long de cette thèse. Je remercie en particulier mon triumvirat parisien qui m'a entourée d'une affection indéfectible et à qui je dois plus que je ne saurais dire : Gwenaël, qui n'a pas son pareil pour trouver les mots justes, quelles que soient les circonstances, et dont la présence au quotidien a rythmé l'écriture de cette thèse, Thierry pour son enthousiasme débordant et Margaux pour sa joie de vivre permanente. Je dois aussi beaucoup à la sagesse de ma précieuse amie Anne-Laure, ainsi qu'à Lila, Tanguy, et bien sûr, à Marie.

Je remercie Fanny, qui a illuminé ces derniers mois de thèse, par sa prévenance et son humour.

Ma famille est un clan soudé, qui ne cesse de m'émerveiller par son énergie et sa force motrice, et tous ont contribué par leurs encouragements et leur affection à rendre cette thèse plus légère. Je suis infiniment reconnaissante à mes parents pour leur amour inconditionnel, leur soutien quotidien et leur phlegme remarquable devant

mes nombreux changements de trajectoire. Je remercie ma sœur et mon frère, pour notre complicité de tous les instants et pour leurs attentions du quotidien. Enfin, je veux dire toute mon admiration à mon grand-père, qui par sa curiosité intellectuelle et son humanité, est un modèle aussi bien pour la philosophe que pour le médecin que je souhaite devenir.

Sommaire

Liste des figures 12

Liste des tableaux 17

Avertissements 18

Liste des abréviations 19

Introduction 21

Partie 1 : Du concept de maladie génétique à la généticisation des maladies 35

Chapitre 1 : Problèmes scientifiques soulevés par le concept de maladie génétique ... 39

Chapitre 2 : Problèmes philosophiques soulevés par le concept de maladie génétique : généticisation, génocentrisme et problème de la sélection causale 97

Partie 2 : De la généticisation des maladies aux conditions de possibilité et aux critères d'une théorie génétique de la maladie 133

Chapitre 3 : Comprendre la généticisation au sein d'une explication interactionniste de la maladie 139

Chapitre 4 : Établir des critères opérationnels pour identifier une théorie génétique 209

Partie 3 : Deux exemples contemporains de théorie génétique 269

Chapitre 5 : La théorie génétique des maladies infectieuses, un exemple de théorie génétique régionale 277

Chapitre 6 : La médecine des réseaux, un cadre pour une théorie génétique de la maladie ? 339

Conclusion 437

Annexes 453

Glossaire 487

Bibliographie 507

Index des auteurs cités 541

Table des matières 547

Liste des figures

Figure 1 : Évolution du nombre d'entrées dans OMIM depuis sa création 54

Figure 2 : Capture d'écran de la page "titre" de l'entrée "phénylcétonurie" sur le site OMIM 57

Figure 3 : Capture d'écran de l'onglet renvoyant aux liens externes accessibles à partir d'OMIM pour la phénylcétonurie 58

Figure 4 : Liste (partielle) des variants alléliques pour le gène de la phénylcétonurie (*612349) 74

Figure 5 : Facteurs influençant le phénotype de la phénylcétonurie 90

Figure 6 : Diagramme de l'action conjointe des facteurs génétiques (G) et environnementaux (E) sur les malades 93

Figure 7 : Nombre d'études d'association pangénomiques publiées entre 2005 et juin 2012 155

Figure 8 : Exemples de maladies pour lesquelles se pose le problème de l'héritabilité manquante 158

Figure 9 : Différents types d'architectures alléliques des maladies 163

Figure 10 : Représentation des phénotypes (exemple : couleur de pelage) obtenus pour deux allèles interagissant de façon épistatique selon la définition de Bateson 169

Figure 11 : Norme de réaction pour l'activité de la MAOA, la maltraitance dans l'enfance et le comportement antisocial 179

Figure 12 : La contribution quantitative des composants de l'héritabilité inclusive 186

Figure 13 : De l'hypothèse scientifique à la théorie scientifique 216

Figure 14 : Représentation schématique de l'organisation hiérarchique des explications des maladies dans la théorie médicale contemporaine avec des exemples de maladies individuelles 225

Figure 15 : Structure causale du concept de maladie 227

Figure 16 : Réseau causal général de l'ulcère duodénal 229

Figure 17 : Comparaison des structures causales de la maladie dans la théorie hippocratique (à gauche) et dans la théorie des germes (à droite) 230

Figure 18 : Quatre critères pour caractériser une théorie médicale 246

Figure 19 : Condition de possibilité, critères d'identification et critères d'évaluation d'une théorie génétique 262

Figure 20 : Passer d'une théorie des maladies génétiques à une théorie génétique générale 264

Figure 21 : Représentation typique d'une théorie génétique des maladies 264

Figure 22 : Représentation typique d'une théorie génétique de la maladie 265

Figure 23 : Représentation schématique du continuum génétique de la maladie infectieuse 304

Figure 24 : Diagramme de l'action conjointe des facteurs génétiques (G) et environnementaux (E) sur les maladies 305

Figure 25 : L'hypothèse d'une architecture génétique des maladies infectieuses 308

Figure 26 : Taux de mortalité annuelle (en échelle algorithmique) comme une fonction de l'âge avant (A) et après (B) le développement de l'hygiène, la vaccination et la chimiothérapie microbienne 309

Figure 27 : Taux de mortalité pour la tuberculose disséminée (bleu) et la tuberculose chronique pulmonaire (rose) pour 100 000 individus non traités d'âge varié en Bavière en 1905 310

Figure 28 : Les quatre théories complémentaires des maladies infectieuses 316

Figure 29 : Représentation schématique des interactions hôte-environnement au cours d'une infection 317

Figure 30 : Contributions respectives de la génétique de l'hôte et de la génétique du microbe dans le développement clinique des maladies infectieuses 320

Figure 31 : Modèle en deux étapes de la susceptibilité génétique à la lèpre 323

Figure 32 : Erreurs innées de l'immunité dépendante de l'axe TLR3-IFN α/β dans l'encéphalite herpétique infantile (HSE) 327

Figure 33 : Réseaux complexes qui sont directement pertinents pour la médecine des réseaux 342

Figure 34 : Représentation de l'approche descendante ("*top-down*") et de l'approche ascendante ("*bottom-up*") dans la biologie des systèmes 345

Figure 35 : Graphe surimposé sur une représentation des sept ponts de Königsberg 357

Figure 36 : Représentation d'un graphe non orienté (A) et d'un graphe orienté (B) 358

Figure 37 : Différences entre un réseau aléatoire et un réseau invariant d'échelle 359

Figure 38 : Représentation des propriétés des réseaux aléatoires, invariants d'échelle et hiérarchiques 362

Figure 39 : Réseau d'interactions protéine-protéine de la levure 364

Figure 40 : Naissance d'un réseau invariant d'échelle par croissance et attachement préférentiel 364

Figure 41 : Origine évolutionnaire et mécanisme de développement d'un réseau invariant d'échelle 366

Figure 42 : Différents types de robustesse à différents types de perturbations pour les réseaux invariants d'échelle et les réseaux invariants aléatoires 368

Figure 43 : Représentation des fondements conceptuels de la médecine des réseaux. 373

Figure 44 : Construction du diseasome 376

Figure 45 : Diseasome - A. Le réseau des maladies humaines. B. Le réseau des gènes de la maladie humaine 377

Figure 46 : Différence de connectivité selon les classes de maladie 381

Figure 47 : Les gènes de la maladie au sein de l'interactome 386

Figure 48 : Représentation schématique de trois concepts de modules de la maladie : le module topologique, le module fonctionnel et le module de la maladie 390

Figure 49 : Les étapes de l'identification et de la validation d'un module de la maladie 392

Figure 50 : Variété des perturbations possibles d'un module fonctionnel de la maladie 394

Figure 51 : Combinaison de quatre réseaux modulaires impliqués dans le développement d'une maladie 397

Figure 52 : Module fonctionnel de la drépanocytose 399

Figure 53 : Comparaison de l'approche pharmacologique conventionnelle et de l'approche pharmacologique fondée sur la médecine des réseaux 402

Figure 54 : Relations possibles entre des gènes et des phénotypes dans une famille de syndromes 405

Figure 55 : Représenter la variété des maladies monogéniques et polygéniques 407

Figure 56 : Trois méthodes pour prédire les gènes candidats d'une maladie : méthode de liaison, méthode fondée sur le module de la maladie et méthode de diffusion 410

Figure 57 : Un réseau écologique dans une communauté 413

Figure 58 : Du génome au réseau d'interactions protéiques du virus, une approche transversale de biologie moléculaire systémique 419

Figure 59 : La cellule infectée virtuelle, un modèle systémique de l'infection 420

Figure 60 : Le réseau de l'infectome humain et le réseau de l'infectome-diseasome humain 422

Figure 61 : Représentation du réseau de l'infectome humain 424

Figure 62 : Décomposition modulaire du réseau de l'infectome humain 425

Figure 63 : Le réseau infectome-diseasome humain (version non étendue) 427

Figure 64 : Le réseau de l'infectome et du diseasome auto-immun 429

Figure 65 : Condition de possibilité, critères d'identification et critères d'évaluation d'une théorie génétique. 441

Figure 66 : Exemples du rôle des ARN non codants dans la pathogénèse 446

Figure 67 : Une vision métagénomique du développement de l'immunité normale et des maladies auto-immunes 448

Figure 68 : Collaboration des modifications épigénétiques et des mutations dans le développement des propriétés des cellules cancéreuses 450

Figure 69 : Les différentes méthodes de "chasse aux gènes" dans les maladies mendéliennes 455

Figure 70 : Composition de la double hélice d'ADN 487

Figure 71 : De l'ADN à l'ARNm et de l'ARNm à la protéine . 489

Figure 72 : Les différents niveaux d'empaquetage de l'ADN dans les chromosomes 490

Figure 73 : Épissage alternatif . 493

Figure 74 : Hétérozygotie et homozygotie, dominance et récessivité 498

Figure 75 : Locus et allèles 499

Figure 76 : Différents types de mutations ont différentes conséquences sur la fonction d'une protéine 501

Figure 77 : Un exemple de polymorphisme génétique : les polymorphismes à un seul nucléotide (SNP) 503

Figure 78 : Comparaison des SNP chez des individus sains et des individus malades 505

Liste des tableaux

Tableau 1 : Liste des principaux codes de la CIM-10 47

Tableau 2 : Liste des principaux codes du chapitre XVII « Malformations congénitales et anomalies chromosomiques » de la CIM-10 49

Tableau 3 : Évolution du système de classification des phénotypes dans OMIM 62

Tableau 4 : Classement par symbole des différents objets référencés sur OMIM 73

Tableau 5 : Présentation d'un exemple fictif de maladie génétique selon Kelly Smith 116

Tableau 6 : La variabilité clinique inter-individuelle est imparfaitement expliquée par la théorie des germes, la théorie microbienne, la théorie immunologique et la théorie écologique des maladies infectieuses. 290

Tableau 7 : Quatre mécanismes génétiques communs dans la théorie génétique des maladies infectieuses 303

Tableau 8 : Cohérence entre l'hypothèse de l'architecture génétique âge-dépendante des maladies infectieuses et les connaissances en immunologie 312

Tableau 9 : La distinction entre maladies monogéniques et maladies polygéniques dans la génétique classique et dans la théorie génétique de la médecine des réseaux 409

Avertissements

1) La plupart des références sont indiquées dans le corps du texte sous un format synthétique (nom de l'auteur, année de publication). Les références complètes sont données dans la bibliographie. J'ai fait le choix de ne mentionner que les initiales du ou des prénoms des auteurs.

2) Lorsque les figures sont reprises d'articles, d'ouvrages ou de sites internet, leur source est indiquée. Les autres figures sont de ma main.

3) Deux figures apparaissent à deux reprises dans des chapitres différents. Elles apparaissent donc deux fois dans la liste des figures sous des numéros différents. Ainsi, la figure 6 est la même que la figure 24 et la figure 19 est la même que la figure 65.

4) Les traductions de texte en langue étrangère sont les miennes.

5) Lorsque le vocabulaire scientifique a été forgé dans une langue étrangère, j'ai essayé de proposer dans la mesure du possible une traduction française. Cependant, lorsque cette traduction n'est pas communément admise et employée, j'ai préféré conserver le terme anglo-saxon original, alors indiqué en italique. Par exemple, pour le terme « *hub protein* », j'ai proposé la traduction de « protéine carrefour » mais dans la suite du texte, nous parlons de protéine *hub*.

6) La liste des abréviations employées est fournie ci-après.

7) Les définitions du glossaire sont librement inspirées de plusieurs ouvrages de biologie moléculaire (Alberts *et al.,* 2013 ; Bassaglia, 2013). La plupart des définitions s'appliquent aux organismes eucaryotes, sauf mention contraire.

8) Dans la suite de ce travail, j'ai adopté l'usage du nous de courtoisie.

Liste des abréviations

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

ARNm : Acide ribonucléique messager

CIM : Classification Internationale des Maladies

DGN : *Disease Gene Network* : Réseau des gènes de la maladie humaine

HbA : Hémoglobine A

HbS : Hémoglobine S

HDN : *Human Disease Network* : Réseau de la maladie humaine

HIN : *Human Infectome Network* : Réseau de l'infectome humain

HIDN : *Human Infectome-Diseasome Network* : Réseau de l'infectome-diseasome humain

OMIM : *Online Mendelian Inheritance on Man* (base de données)

OMS : Organisation Mondiale de la Santé

TGMI : Théorie génétique des maladies infectieuses

VCI : Variabilité clinique interindividuelle

Introduction

Le paradoxe de la génétique médicale contemporaine

La phénylcétonurie, dont les symptômes ont été décrits dès 1934 par le médecin norvégien Asbjørn Følling, est une maladie métabolique liée au déficit d'une enzyme hépatique, la phénylalanine hydroxylase. La phénylalanine hydroxylase permet de transformer la phénylalanine, un acide aminé que l'on trouve dans l'alimentation, en un autre acide aminé, la tyrosine. Lorsque cette enzyme est mutée, la phénylalanine ne peut être transformée en tyrosine et s'accumule dans le sang, exerçant alors un effet toxique sur le système nerveux central. Non traitée, la phénylcétonurie est donc responsable d'un retard mental. Cependant, un simple régime sans phénylalanine, s'il est bien conduit dès la naissance, prévient l'apparition des symptômes de la maladie. Lorsqu'en 1963, Robert Guthrie met au point un test biochimique permettant de détecter cette maladie sur simple prise de sang, un dépistage systématique de la phénylcétonurie chez les nouveau-nés est rapidement mis en place dans de nombreux pays occidentaux, dont les États-Unis et la France. La phénylcétonurie est une maladie héréditaire, qui a une explication moléculaire, un test de dépistage et un traitement : elle devient une maladie-star de la génétique humaine.

La phénylcétonurie incarne alors un exemple paradigmatique du concept de maladie génétique tel qu'il est pensé dans les années 1960 : une maladie rare, héréditaire, mendélienne et monogénique. En effet, la phénylcétonurie est une maladie rare dont la prévalence varie de 1/4 000 à 1/40 000. C'est une maladie héréditaire, c'est-à-dire transmise par les parents. C'est une maladie mendélienne, transmise sur le mode autosomique récessif, ce qui signifie qu'il faut deux allèles mutés pour que la maladie apparaisse. C'est une maladie monogénique, c'est-à-dire qui est causée par la mutation d'un seul gène, le gène PAH qui sera cloné dans les années 1980 et qui code l'enzyme phénylalanine hydroxylase. Le modèle de la phénylcétonurie est donc celui d'une mutation héréditaire et mendélienne d'un gène, entraînant une protéine mutée déficiente, qui cause à son tour la maladie. Cette définition de la maladie génétique s'accorde parfaitement avec le dogme central de la

biologie moléculaire et avec le concept moléculaire du gène, qui se développent à la même époque. Le dogme central de la biologie moléculaire est l'idée selon laquelle il existe une correspondance linéaire et spécifique entre une séquence de nucléotides et les acides aminés qu'ils codent. Le concept moléculaire du gène identifie la structure d'un gène à sa fonction, le codage de la séquence d'acides aminés constituant une protéine.

Cependant, depuis l'établissement de la phénylcétonurie comme exemple paradigmatique de la maladie génétique, un double mouvement s'est opéré au sein de la génétique médicale.

D'une part, le concept de maladie génétique s'est étendu, bien au-delà du concept de maladie monogénique, dont il était jusqu'ici synonyme. Plusieurs découvertes ont contribué à cet élargissement du concept de maladie génétique. La découverte des gènes de susceptibilité dans les années 1970, (gènes qui sont associés à l'apparition d'une maladie mais qui ne suffisent pas à la causer), et la mise en évidence des oncogènes et des anti-oncogènes dans les années 1980 (gènes dont l'activation ou l'inactivation est décisive dans le développement du cancer), ont attiré l'attention sur la génétique de maladies polygéniques communes. L'essor des techniques de séquençage de l'ADN, et plus généralement des techniques de génie génétique, a permis de mettre au point des méthodes d'identification des variants alléliques. Dans la littérature biomédicale contemporaine, toute maladie dont l'apparition est statistiquement associée à un variant allélique, c'est-à-dire à une variation d'un ou plusieurs nucléotides dans un gène, tend à être considérée comme génétique. Le concept de maladie génétique s'applique donc aujourd'hui à des maladies communes. Ces maladies ne sont pas héréditaires, mais dues à des mutations *de novo* (apparues au moment de la formation du zygote) ou à des mutations acquises (sous l'effet de perturbations environnementales, par exemple). Leur transmission ne suit pas les lois de Mendel, et elles sont dites polygéniques ou complexes, parce que leur physiopathologie implique l'action conjointe de plusieurs gènes et de facteurs environnementaux. Ainsi, le cancer, le diabète, l'hypertension artérielle, et même la tuberculose – maladie environnementale par excellence puisque causée par un agent infectieux –, ont progressivement été considérées comme des maladies génétiques.

D'autre part, une série de découvertes scientifiques a bouleversé notre compréhension de la maladie monogénique et a contribué à dissoudre la distinction forte établie entre maladie monogénique simple et maladie polygénique complexe. Ainsi, la physiopathologie de la phénylcétonurie a révélé plusieurs surprises. Des phénomènes d'hétérogénéité allélique (plus de 500 mutations du gènes PAH peuvent causer la phénylcétonurie), d'hétérogénéité génétique (en l'absence de mutation du gène PAH, une mutation du gène BH4 qui code son récepteur peut suffire à causer la maladie) et l'existence de gènes modificateurs (le gène BH4 influence l'expression du gène PAH et les conséquences des mutations de ce gène sur la sévérité et la forme des symptômes) ont remis en cause la correspondance linéaire entre l'existence d'une mutation dans le gène PAH, la production d'une protéine PAH mutée et l'apparition des symptômes de la maladie. Ces phénomènes d'hétérogénéité allélique, d'hétérogénéité génétique et de gènes modificateurs ont remis en cause l'idée selon laquelle les maladies monogéniques étaient des maladies simples. En parallèle, la découverte de l'épissage alternatif (un même ARN messager peut donner lieu à plusieurs ARN transcrits matures), de l'existence d'éléments transposables dans le génome, de séquences d'ADN régulatrices situées à plusieurs milliers de base de la séquence d'ADN qu'elles contrôlent, bouleversent le concept moléculaire du gène et le dogme central de la biologie moléculaire. Ce concept de gène morcelé qui donne une image plus complexe de la relation entre génotype et phénotype, ne fait que renforcer la remise en cause du concept de maladie monogénique.

Enjeux du paradoxe de la génétique médicale

La génétique médicale est donc aujourd'hui au cœur d'un paradoxe. D'une part, le concept de maladie génétique s'est étendu bien au-delà du modèle de la maladie monogénique, au point de pouvoir concerner toute maladie. D'autre part, le modèle de la maladie monogénique simple a volé en éclats et la distinction entre maladies monogéniques et maladies polygéniques est devenue de plus en plus difficile à établir, au point qu'il n'y a pas de consensus sur la définition du concept de maladie génétique. L'enjeu de ce paradoxe est double, à la fois épistémologique et éthique.

Sur le plan épistémologique, il s'agit à la fois d'interroger la légitimité d'une distinction entre maladie génétique et non génétique et de savoir s'il est possible de parvenir à unifier l'ensemble des connaissances disparates de la génétique des maladies. Cette question est d'autant plus pressante que depuis l'achèvement du Projet Génome Humain, les dix dernières années, aussi appelées « ère post-génomique », ont constitué une période de développement exponentiel du nombre et de la diversité de données acquises en génétique humaine. Ainsi, si le séquençage du premier génome humain a pris plus de dix ans et coûté trois milliards de dollars, la compagnie Illumina a annoncé en janvier 2014 que sa dernière technologie permettait de séquencer un génome entier, chaque paire de bases étant lue trente fois par la machine, en trois jours, pour un prix de mille dollars. De nouvelles méthodes d'identification des variants alléliques dans les maladies complexes ont été mises au point comme les études d'association pangénomiques, qui permettent de comparer les génomes de plusieurs milliers d'individus. Les données ne concernent plus seulement la génomique (l'étude fonctionnelle du génome), mais aussi les interactions entre le génome, le transcriptome (ensemble des différents types d'ARN transcrits à partir d'un génome), le protéome (ensemble des protéines traduites à partir d'un génome). Interpréter ce déluge de données de l'ère post-génomique et leur donner un sens dans le domaine de la génétique médicale est devenue une tâche de plus en plus difficile.

Sur le plan éthique, il s'agit de déterminer à quel point l'exceptionnalisme génétique est justifié et quel serait l'impact éthique, juridique et social d'une extension du concept de maladie génétique. L'exceptionnalisme génétique consiste à considérer que les maladies génétiques doivent faire l'objet d'un traitement éthique particulier, parce qu'elles posent des problèmes éthiques spécifiques et distincts des maladies non génétiques. C'est l'exceptionnalisme génétique qui permet de justifier une législation spécifique pour encadrer le dépistage, le diagnostic et le traitement des maladies génétiques. Si l'on admet qu'il n'y a pas de différence épistémologique entre maladie génétique et maladie non génétique, peut-on encore soutenir l'exceptionnalisme génétique ? Plus généralement, du diagnostic à la thérapeutique, de la représentation individuelle de la maladie vécue comme une forme de « destin génétique » aux questions juridico-sociales comme le secret médical génétique et les assurances-

maladies ou comme la prolifération des tests génétiques, quel serait l'impact d'un reclassement de toute maladie comme génétique ?

Une résolution traditionnelle insatisfaisante du paradoxe de la génétique médicale

La philosophie de la médecine est donc doublement concernée par la résolution de ce paradoxe de la génétique médicale contemporaine. Nous ne traiterons cependant que l'angle épistémologique de ce paradoxe, à la fois parce que traiter simultanément ces deux enjeux dépasserait les limites de ce travail et parce qu'il nous semble que la résolution du problème épistémologique est déterminante pour aborder le problème éthique.

Cette thèse est donc un travail d'épistémologie des sciences biomédicales dont l'objectif général est de comprendre ce paradoxe, selon lequel il n'existe pas de définition consensuelle et univoque du concept de maladie génétique, mais où toute maladie est considérée comme génétique.

D'aucuns pourraient soutenir que ce paradoxe semble avoir déjà été résolu : les sociologues comme les philosophes considèrent comme illégitime cette extension du concept de maladie génétique. Les sociologues arguent que cette extension relève au mieux d'une stratégie de levée de fonds en recherche biomédicale et au pire d'un essentialisme génétique qui tend à attribuer toute différence entre deux individus au seul pouvoir du gène. Ils appellent « généticisation » cette extension du concept de maladie génétique, en donnant à ce terme une connotation négative intrinsèque. Les philosophes dénoncent cette extension au mieux comme un truisme (les gènes sont impliqués dans tout phénomène biologique d'une façon ou d'une autre) et au pire comme une survivance du génocentrisme qui tend à attribuer une priorité causale aux gènes dans tous les phénomènes biologiques. En quoi cette résolution du paradoxe ne nous paraît-elle pas satisfaisante ?

Les philosophes qui refusent l'extension du concept de maladie génétique au nom d'une position anti-génocentriste, cherchent simultanément à redéfinir le concept de maladie génétique de façon plus stricte. Parvenir à redéfinir le concept de

maladie génétique équivaut alors à délimiter quelle est l'extension du concept de maladie génétique en tant que classe de maladie, distincte de la classe des maladies non génétiques. Il s'agit de déterminer pour quelles maladies on peut résoudre le problème de la sélection causale en faveur des gènes. Autrement dit, il s'agit de déterminer pour quelles maladies on peut sélectionner le gène comme la cause la plus importante, quelle que soit la stratégie utilisée pour justifier le sens de cette priorité causale du gène sur les autres facteurs déterminants d'une maladie.

Cette approche de la résolution du paradoxe ne nous satisfait pas pour trois raisons. Tout d'abord, aucune des stratégies que nous examinerons ne parvient à redéfinir un concept de maladie génétique au sens strict, qui permette de distinguer de façon convaincante maladies génétiques et maladies non génétiques. Surtout, cette approche est un retour déguisé au génocentrisme qu'elle prétend combattre. Il s'agit encore une fois d'affirmer que les gènes sont la cause la plus importante *au moins pour certaines maladies*. Enfin cette approche ne permet pas de rendre compte de façon unifiée du rôle des gènes à la fois dans les maladies dites monogéniques simples et dans les maladies dites polygéniques complexes.

La première thèse que nous défendons dans ce travail est qu'une résolution cohérente du paradoxe de la génétique médicale contemporaine nécessite d'adopter la stratégie inverse à celle qui est traditionnellement proposée. Au lieu de chercher à établir des critères pour distinguer maladies génétiques et non génétiques, il faut abandonner le concept de maladie génétique. Au lieu de condamner la généticisation des maladies comme un projet génocentriste, il faut redonner à ce terme un sens axiologiquement neutre (que certains sociologues défendent d'ailleurs), et chercher à comprendre en quel sens, ni trivial, ni génocentriste, toute maladie peut être considérée comme génétique.

L'hypothèse d'une théorie génétique de la maladie

Le contexte scientifique et technologique actuel de la génétique médicale est marqué par l'identification toujours plus rapide d'associations toujours plus nombreuses, mais isolées entre des variants alléliques et des maladies. Pour résoudre le paradoxe de la génétique médicale, nous faisons l'hypothèse que la généticisation

des maladies constitue une réorganisation épistémique des connaissances disparates de la génétique médicale. Plus précisément, notre hypothèse est que la généticisation des maladies révèle l'élaboration d'une explication ni triviale ni génocentriste du rôle causal commun des gènes dans toutes les maladies au sein de la littérature biomédicale contemporaine, que nous appelons une « théorie génétique de la maladie ».

Identifier les conditions de possibilité d'une théorie génétique de la maladie

Pour tester cette hypothèse, il nous faut déterminer à quelles conditions la généticisation des maladies peut révéler une théorie génétique de la maladie[1]. Autrement dit, il s'agit de déterminer si une théorie génétique de la maladie peut exister en principe et à quelles conditions elle peut exister. Or, le concept de théorie génétique de la maladie pose une double difficulté : une difficulté quant à ce qu'on entend par « génétique » et une difficulté quant à ce qu'on entend par « théorie ».

La première difficulté tient à ce qu'il nous faut déterminer à quelles conditions une théorie génétique de la maladie ne serait ni triviale ni génocentriste. En effet, puisque nous avons établi que la généticisation des maladies ne saurait ni se réduire à l'affirmation triviale que les gènes jouent un rôle quelconque dans toutes les maladies, ni se réduire à l'affirmation génocentriste que les gènes sont la cause principale de toutes les maladies, la généticisation des maladies comme élaboration d'une théorie génétique de la maladie ne peut faire sens qu'à condition qu'on établisse un cadre interactionniste pour l'explication de toutes les maladies. Nous démontrerons que l'explication interactionniste la plus adaptée est une explication interactionniste de type co-constructionniste, qui permet une description complexe de la matrice causale d'une maladie donnée, tout en permettant de distinguer d'un point de vue heuristique le rôle des gènes du rôle des autres facteurs explicatifs d'une maladie.

La deuxième difficulté tient au fait que le terme de théorie est un concept hautement polémique et polysémique en philosophie des sciences. Si le terme de

[1] Jusqu'au chapitre 4, nous employons le terme de « théorie génétique de la maladie » au sens d'explication du rôle commun des gènes dans toutes les maladies. A partir du chapitre 4, nous définissons deux formes de théories génétiques particulières et distinguons *la théorie génétique de la maladie* de la *théorie génétique des maladies*.

« théorie » a d'ailleurs été utilisé à plusieurs reprises dans l'histoire de la médecine (la théorie des humeurs, la théorie des germes, etc.), seuls quatre philosophes contemporains majeurs se sont intéressés à la structure des théories biomédicales : Kenneth Schaffner, Paul Thompson, Kazem Sadegh-Zadeh et Paul Thagard[2]. Étant donné notre objectif, nous ne pouvons cependant reprendre aucune de ces conceptions entièrement à notre compte. En effet, aucune de ces conceptions ne permet simultanément de caractériser spécifiquement les théories médicales en les différenciant des théories biologiques, de rendre compte de la variété des niveaux d'explication en médecine (selon qu'on s'intéresse à l'explication d'une maladie, d'une classe de maladie ou de la maladie en général) et du pluralisme explicatif à l'œuvre en médecine (explication physiopathologique, évolutionnaire ou épidémiologique) et de définir des critères opérationnels qui permettent d'identifier une théorie médicale au sein de la littérature biomédicale contemporaine.

Nous proposons donc une définition originale du concept de théorie médicale, que nous qualifions de minimalement normative, au sens où elle est fondée sur une conception peu exigeante et déflationniste des théories scientifiques et où elle permet de définir *a minima* les critères nécessaires d'une théorie médicale. Une théorie médicale *a minima suppose au minimum un type d'explication (évolutionnaire, épidémiologique, ou physiopathologique) d'au minimum une cible explicative (causes, symptômes, évolution, ou traitement) d'une maladie, ou d'une classe de maladie ou de la maladie en général*. En combinant ces différents critères, on définit le spectre des théories médicales possibles.

À partir de cette définition du concept de théorie médicale, il est possible de définir une théorie génétique *a minima*, comme une théorie médicale partielle qui se concentre sur le rôle commun des gènes dans l'explication d'une maladie, d'une classe de maladie ou de toutes les maladies. Et à partir de cette définition du concept de théorie génétique *a minima*, on peut envisager un spectre de théories génétiques possibles, selon le niveau d'explication, le type de cibles explicatives et le type d'explication mobilisés. Parce que notre hypothèse générale est de comprendre la généticisation non pas d'une maladie ou d'une classe de maladie, mais de l'ensemble

[2] Nous revenons sur ce point au chapitre 4.

des maladies, nous distinguons deux formes extrêmes de théorie génétique générale, parmi le spectre des théories génétiques possibles : une *théorie génétique des maladies* et une *théorie génétique de la maladie*. Ces deux formes extrêmes de théorie génétique générale ont également pour objectif d'expliquer le rôle central des gènes dans toutes les maladies, mais elles n'ont pas la même structure. Une *théorie génétique des maladies* est une somme de théories génétiques régionales, chaque classe de maladie ayant sa propre théorie génétique : la théorie génétique des maladies infectieuses, la théorie génétique des maladies cardiovasculaires, la théorie génétique des maladies auto-immunes, etc. Au contraire, une *théorie génétique de la maladie* repose sur une définition commune du rôle des gènes dans la maladie en général, à partir de laquelle chaque maladie peut être expliquée du point de vue génétique. Ces deux formes différentes de théories générales génétiques n'ont donc pas le même type de conséquences sur notre compréhension du rôle commun des gènes dans l'explication des maladies. Par exemple, une *théorie génétique des maladies* respecte notre classification des maladies actuelles, tandis qu'une *théorie génétique de la maladie* n'est pas structurée en fonction des classes de maladie que nous connaissons.

C'est pourquoi nous proposons de raffiner notre hypothèse initiale en une alternative : ou bien la généticisation révèle *une théorie génétique des maladies*, c'est-à-dire un ensemble hétérogène de théories génétiques régionales, chacune étant spécifique d'une classe de maladie, ou bien la généticisation révèle *une théorie génétique de la maladie*, c'est-à-dire une théorie qui repose sur une définition de la maladie qui unifie le rôle commun des gènes, à partir de laquelle toute maladie pourrait être instanciée et qui serait susceptible de renouveler en profondeur la façon dont nous classons les maladies.

Tester notre alternative

Mais comment tester cette alternative ? Cette question pose un problème méthodologique, mais interroge aussi la légitimité du philosophe, lorsqu'il s'agit de déterminer si une théorie scientifique existe. N'est-ce pas plutôt la responsabilité du scientifique ?

En aucun cas, nous ne souhaitons nous substituer au travail du scientifique : il ne s'agit pas de formuler nous-mêmes le contenu d'une hypothétique théorie génétique de la maladie. Nous n'avons pas non plus l'ambition de nous substituer au sociologue et de faire un travail de terrain, qui permettrait d'identifier des théories génétiques implicites dans le discours des scientifiques. Enfin, tester cette alternative ne nécessite pas de recenser exhaustivement l'ensemble des théories génétiques, éventuellement en concurrence, qui émergent aujourd'hui dans le domaine de la génétique des maladies humaines. Puisque nous montrons qu'une théorie génétique est possible en principe et à quelles conditions elle est possible, il s'agit de tester si la généticisation révèle plutôt l'élaboration d'une *théorie génétique des maladies* ou d'une *théorie génétique de la maladie* et quelles sont les conséquences de ces deux formes de théories génétiques générales sur notre compréhension des maladies. Nous confronterons donc les critères que nous avons définis pour chacune de ces hypothèses à des exemples issus de la littérature biomédicale.

Tester l'hypothèse d'une théorie génétique des maladies semble supposer qu'on établisse s'il existe une théorie génétique régionale pour chaque classe de maladie connue. Ce travail aurait très largement débordé le cadre de notre thèse, aussi avons-nous choisi de prendre l'exemple d'une théorie génétique explicitement revendiquée comme telle, qui concerne une classe de maladie et que nous utilisons comme approximation pour penser la structure d'une théorie génétique des maladies. C'est l'exemple de la théorie génétique des maladies infectieuses, défendue par l'équipe de recherche du Laboratoire de Génétique des Maladies Infectieuses de l'Hôpital Necker.

Tester l'hypothèse d'une théorie génétique de la maladie s'avère plus difficile, puisqu'aucun biologiste ou chercheur en génétique médicale ne revendique à l'heure actuelle une *théorie génétique de la maladie*. Nous nous sommes donc tournée vers des programmes de recherche qui cherchent à comprendre les principes d'organisation généraux des phénomènes biologiques et des phénomènes physiopathologiques, comme la biologie et la médecine des systèmes. Dans ce champ de recherche très récent, nous identifions la médecine des réseaux comme une théorie générale de la maladie, au sein de laquelle on peut isoler une théorie génétique de la maladie en cours d'élaboration.

Thèses soutenues et plan de l'argumentation

L'objectif de notre thèse est de comprendre le paradoxe de la génétique médicale contemporaine, où il n'a jamais été aussi difficile de définir le concept de maladie génétique mais où toute maladie est progressivement considérée comme génétique. Nous défendrons trois thèses :

(1) Les tentatives de résolution traditionnelle de ce paradoxe sont contradictoires. La seule façon de résoudre ce paradoxe consiste à abandonner le concept de maladie génétique et à prendre au sérieux la généticisation des maladies.

(2) Avant l'examen de tout exemple issu de la littérature biomédicale contemporaine, il faut définir les conditions de possibilité d'une théorie génétique afin que celle-ci ne soit ni génocentriste ni triviale.

(2a) La première de ces conditions est que pour éviter à la fois le génocentrisme et la trivialité, une théorie génétique ne peut exister qu'au sein d'une explication interactionniste co-constructionniste des maladies. Nous définissons cet interactionnisme co-constructionniste comme une explication des maladies qui prend en compte les interactions réciproques entre les gènes et les autres facteurs déterminants d'une maladie, sans donner de prédominance causale à aucun de ces facteurs.

(2b) La seconde de ces conditions est qu'une théorie génétique *a minima* est une théorie partielle, centrée sur le rôle causal commun des gènes, qui explique au minimum une cible explicative (causes, symptômes, évolution, traitement) au niveau soit d'une maladie, d'une classe de maladie, ou de la maladie en général, en mobilisant au moins un type d'explication (physiopathologique,

évolutionnaire, épidémiologique). À partir d'une combinaison de ces critères, on peut potentiellement définir un spectre extrêmement varié de théories génétiques, que ce soit en termes de contenu, de niveau d'explication ou de type d'explication.

(3) En particulier, coexistent aujourd'hui dans la littérature biomédicale contemporaine au moins un exemple de théorie génétique d'une classe de maladie et un exemple de théorie génétique de la maladie en général. Ces deux exemples de théories génétiques ont des limites (expérimentales et conceptuelles), mais elles renouvellent à des degrés divers notre compréhension de la génétique des maladies.

Le plan de notre argumentation correspond à ces trois thèses. Dans la première partie, il s'agit de proposer une vision synthétique des problèmes scientifiques et philosophiques que posent à la fois le bouleversement du concept de maladie génétique et la généticisation des maladies. Le chapitre 1 est consacré à l'examen des problèmes scientifiques, en passant par l'analyse de l'évolution de la base de données OMIM, fondée en 1966 et qui est la classification internationale de référence des maladies génétiques. Le chapitre 2 est consacré à l'examen des tentatives traditionnelles de résolution du paradoxe de la génétique médicale contemporaine. En constatant l'échec de ces tentatives, nous concluons à la nécessité d'abandonner le concept de maladie génétique et posons l'hypothèse que la généticisation révèle l'élaboration d'une explication du rôle commun des gènes dans toutes les maladies, que nous appelons « théorie génétique de la maladie ».

Dans la seconde partie, nous définissons les conditions de possibilité d'une théorie génétique de la maladie, avant examen de tout exemple de théorie génétique issu de la littérature biomédicale contemporaine. Le chapitre 3 examine deux concepts d'interactionnisme, l'interactionnisme de la génétique quantitative centré autour du concept d'héritabilité et l'interactionnisme co-constructionniste élaboré par les biologistes du développement. Le chapitre 4 est consacré à la définition des conditions de possibilité d'une théorie génétique, en passant par l'analyse du concept de théorie

médicale. Cette deuxième partie nous permet de raffiner notre hypothèse en une alternative : la généticisation des maladies révèle soit une théorie génétique des maladies, ensemble hétérogène de théories génétiques régionales, chaque théorie génétique étant spécifique d'une classe de maladie, soit une théorie génétique de la maladie, fondée sur une définition du concept de maladie qui spécifie le rôle commun des gènes et qui peut être instanciée pour chaque maladie individuelle, renouvelant potentiellement la façon dont nous classons les maladies.

La troisième partie consiste à confronter les critères développés dans la deuxième partie à deux exemples de théories génétiques issus de la littérature biomédicale contemporaine. Le chapitre 5 est consacré à l'examen de la théorie génétique des maladies infectieuses et utilise cette théorie génétique pour penser à quoi ressemblerait une théorie génétique des maladies. Le chapitre 6 est consacré à l'examen de la médecine des réseaux comme théorie générale de la maladie au sein de laquelle on peut identifier une théorie génétique de la maladie.

Enfin, en conclusion, nous reviendrons sur cette coexistence entre théorie génétique des maladies et théorie génétique de la maladie dans la structure du savoir biomédical contemporain. Puisque nous montrons qu'il existe une multiplicité de théories génétiques en principe et de fait, il s'agit de déterminer si ces théories génétiques sont en concurrence ou si au contraire elles sont dans un rapport de complémentarité dans l'explication du rôle commun des gènes dans toutes les maladies.

Partie 1 : Du concept de maladie génétique à la généticisation des maladies

Cette première partie propose une vision synthétique des problèmes contemporains que soulève le concept de maladie génétique, aussi bien dans sa définition scientifique que dans sa définition philosophique.

Dans le chapitre 1, nous commençons par aborder les difficultés d'une définition scientifique de la maladie génétique, en choisissant de présenter l'évolution du concept de maladie génétique à travers l'histoire des classifications dont les maladies génétiques ont fait l'objet. Il n'existe cependant aucune classification « officielle » des maladies génétiques, qui fasse l'objet d'une reconnaissance institutionnelle scientifique mondiale. Aussi, après avoir montré que la Classification Internationale des Maladies (la CIM-10), du fait même de ses objectifs et de son histoire, ne peut servir de support à une classification des maladies génétiques, nous nous tournons vers l'analyse de la base de données *Online Mendelian Inheritance On Man (OMIM)*. Créé en 1967 par Victor McKusick dans une version papier, le catalogue des désordres mendéliens est aujourd'hui accessible en ligne et jouit d'une reconnaissance scientifique qui lui confère le statut de classification officieuse des maladies génétiques. Il nous est alors possible d'identifier deux changements majeurs dans OMIM depuis sa création. D'une part, la base de données s'est ouverte progressivement aux maladies de transmission non mendélienne, tandis que l'hérédité mendélienne faisait elle-même l'objet d'une remise en question profonde de ses concepts les plus fondateurs. D'autre part, la base de données n'a plus pour seuls objectifs la classification et la nomenclature des maladies, elle liste également tout variant allélique impliqué dans une maladie donnée. Nous affirmons donc que le concept de maladie génétique est aujourd'hui en crise : il s'est étendu jusqu'à pouvoir concerner virtuellement toute maladie au point qu'on parle de « continuum génétique de la maladie » et de « généticisation des maladies », mais en étendant ainsi son champ de compréhension, il a perdu en clarté conceptuelle.

Dans le chapitre 2, nous nous tournons vers l'interprétation philosophique de ces deux problèmes intrinsèquement liés que sont la définition du concept de maladie génétique et la généticisation des maladies. En reprenant la littérature consacrée à ce sujet, nous identifions une formulation commune de ces deux questions sous l'angle du problème de la sélection causale. Identifier une maladie comme génétique, ce serait choisir de donner une importance causale majeure aux facteurs génétiques au détriment des facteurs environnementaux dans l'explication de la maladie. Généticiser les maladies reviendrait donc à considérer que les gènes sont la cause la plus importante dans l'explication de toutes les maladies. À partir du moment où ces deux problèmes sont formulés en ces termes, les philosophes tendent à adopter une même stratégie : s'opposant à la généticisation, qui est considérée comme une extension abusive du génocentrisme qui tend à donner la priorité causale aux gènes dans l'explication de tous les phénomènes biologiques, ils cherchent pourtant à « sauver » le concept de maladie génétique. Il s'agit alors de trouver un fondement légitime à une résolution du problème de la sélection causale en faveur des gènes dans l'explication d'un certain nombre de maladies – celles que l'on peut identifier raisonnablement comme génétiques. Si le diagnostic est commun, les stratégies diffèrent pour résoudre ce problème de la sélection causale. Nous identifions trois stratégies principales : (1) une stratégie que nous appelons l'approche « cause/condition » et dont l'objectif est de définir les maladies génétiques comme celles où les facteurs génétiques peuvent être considérés comme la cause de la maladie tandis que les facteurs environnementaux sont relégués au rang de simples conditions, (2) une stratégie que nous appelons « différentialiste » et qui consiste à définir les maladies génétiques relativement à une population de contraste donnée où les différences génétiques sont la cause des différences phénotypiques (3) et enfin une stratégie que nous appelons « pragmatique » et qui revient à admettre que le concept de maladie génétique ne peut être défini en soi, même si des explications génétiques des maladies peuvent avoir une pertinence en fonction des contextes et des intérêts des différents acteurs de ces explications. Nous montrons alors que chacune de ces stratégies échoue à définir un concept de maladie génétique satisfaisant, en partie faute de pouvoir s'affranchir du cadre initial du problème de la sélection causale. Nous proposons alors

une stratégie inverse à celle employée par la majorité des philosophes : plutôt que d'abandonner la généticisation et de tenter de sauver le concept de maladie génétique en cherchant une solution au problème de la sélection causale, nous proposons d'abandonner le concept de maladie génétique et le problème de la sélection causale et de faire l'hypothèse que la généticisation des maladies génétiques est le reflet de l'élaboration d'une explication interactionniste du rôle causal que jouent les gènes dans toutes les maladies.

Chapitre 1 : Problèmes scientifiques soulevés par le concept de maladie génétique

1.1. Le concept de maladie génétique au prisme des classifications

1.1.1. Etudier l'histoire d'un concept à travers l'histoire de sa classification

Au cours du vingtième siècle, le concept de maladie génétique a évolué (Yoxen, 1982 ; Lindee, 2000, 2002 ; Melendro-Oliver, 2004) de façon spectaculaire comme le résume Susan Lindee :

> « La maladie héréditaire humaine a une longue histoire derrière elle, mais pendant la plus grande partie du vingtième siècle, les généticiens connaissaient davantage les pathologies héréditaires de la souris ou de la mouche que celles des humains. Pendant les trente dernières années, cependant, la pathologie en tant qu'elle est liée à nos gènes, est devenu un centre d'intérêt majeur des médecins, des scientifiques mais aussi des entreprises. Plusieurs ensembles de symptômes compliqués ont été reconfigurés comme des maladies génétiques et à partir des années 1990, l'hérédité est devenue l'explication qui prévaut pour virtuellement n'importe quel état pathologique. Même les maladies infectieuses ont été subsumées sur le plan de la rhétorique comme l'expression d'une vulnérabilité inhérente par des mécanismes génétiques et comme des réponses du système immunitaire génétiquement déterminées. Le modèle dans lequel tous les symptômes physiques ont une cause génétique sous-jacente est devenu le modèle dominant de la maladie dans le monde industrialisé, en particulier dans les pays anglophones et encore plus aux États-Unis. » (Lindee, 2000, p. 236)

Le concept de maladie génétique ne recouvre donc plus aujourd'hui seulement des maladies rares, héréditaires, mendéliennes, monogéniques comme la phénylcétonurie, l'anémie falciforme ou la mucoviscidose. Il désigne aussi des maladies communes, acquises, non mendéliennes, polygéniques comme le diabète, le cancer ou même les maladies infectieuses.

Pour comprendre la signification du concept de maladie génétique aujourd'hui et les problèmes de définition qu'il pose à la communauté scientifique, il est donc nécessaire de revenir sur l'histoire de ce concept. Mais cette histoire du concept de maladie génétique ne peut se comprendre qu'au prisme de plusieurs histoires. Il faudrait d'abord évoquer l'histoire de la façon dont la génétique médicale s'est établie comme discipline au sens à la fois théorique (établissement des fondements théoriques, découvertes scientifiques et techniques décisives) mais aussi au sens académique et institutionnel du terme (établissement d'un corpus d'ouvrages et d'articles, transmission et enseignement de ce « nouveau » savoir, interactions entre les chercheurs, les laboratoires de recherche, les institutions, les sociétés savantes)[3]. Il faudrait bien sûr revenir sur l'histoire du concept de gène, dont la définition soulève de nombreux débats aujourd'hui[4]. Une histoire des maladies génétiques devrait aussi faire une place à l'histoire des idéologies[5], mais aussi à l'histoire des mouvements sociaux (associations de patients par exemple) et des débats de société (« ELSI » : *Ethical, Legal and Social Issues* autour du Projet Génome Humain) qui ont généré, accompagné ou retardé le développement de la génétique médicale et de ses applications. Enfin, il y aurait une place pour l'histoire des discours scientifiques et non-scientifiques (discours médiatiques et politiques) qui ont été tenus au sujet des maladies génétiques.

Par ailleurs, d'un point de vue méthodologique, cette histoire est d'autant plus complexe qu'elle dépend aussi des cultures locales médicales, scientifiques, techniques, institutionnelles et culturelles. Le développement de la génétique entre la première et la deuxième guerre mondiale n'a pas eu lieu de la même façon et avec la

[3] Pour une vision récente et globale de l'histoire de la génétique médicale, on pourra se reporter à deux ouvrages de Harper (Harper, 2004, 2008). L'ouvrage de 2004 propose un commentaire de nombreux articles de génétique médicale considérés comme essentiels à son histoire, l'ouvrage de 2008 propose une histoire générale de la génétique médicale, qui s'étend de la période pré-mendélienne à la complétion du Projet Génome Humain.

[4] Il est difficile de se limiter à quelques références pour un sujet aussi débattu mais on pourra se référer à l'ouvrage d'Evelyn Fox Keller (Fox Keller, 2003) ou à l'ouvrage collectif édité par Peter Beurton, Raphaël Falk et Hans-Jörg Rheinberger (Beurton *et al.,* 2000).

[5] Pour une vision des liens historiques entre génétique médicale et eugénisme, se reporter par exemple à l'ouvrage « L'éternel retour de l'eugénisme » (Gayon *et al.*, 2006).

même intensité en France, en Angleterre, aux États-Unis et en Allemagne par exemple[6].

Enfin, l'histoire de la notion de maladie génétique est encore compliquée par le fait que certaines maladies dites génétiques ont fait l'objet d'une abondante communication scientifique et d'une vulgarisation médiatique importante, soit parce qu'elles ont fait l'objet de découvertes fracassantes (l'anémie falciforme par exemple est la première maladie génétique dont la structure moléculaire a été identifiée par Neels et Pauling en 1948 et en 1949 (Strasser, 2002)), soit parce qu'elles ont été à l'origine de programmes de dépistage de grande ampleur (le syndrome de Down en France ou la phénylcétonurie aux États-Unis dans les années 1960 (Paul, 1994)) ou bien encore parce qu'elles sont considérées comme un problème de santé publique prioritaire (comme le cancer du sein et les gènes BRCA1 et BRCA2). Ces quelques « maladies-stars » ont souvent fait l'objet d'analyses détaillées[7], mais ne sont pas toujours représentatives de ce qu'on appelle aujourd'hui « maladies génétiques ».

C'est pourquoi, plutôt que de prétendre donner une vision exhaustive de cette histoire du concept de maladie génétique, nous avons choisi de mettre l'accent sur les principales transformations qu'a subi le concept de maladie génétique à travers l'histoire des classifications des maladies génétiques. Pour justifier cette analyse des classifications des maladies génétiques, nous nous appuyons sur trois considérations que Geoffrey Bowker et Susan Leigh Star ont développées dans leur ouvrage *Sorting things out : classification and its consequences* (Bowker et Star, 2000). La première est la reconnaissance de l'impact des classifications sur la vie quotidienne des individus singuliers, un impact particulièrement décisif dans le cadre des politiques publiques de la santé et des classifications en médecine. Ainsi, être désigné comme atteint d'une maladie génétique a un impact sur la vie des patients et il est très différent d'être labellisé porteur d'une maladie génétique et porteur d'une maladie infectieuse.

[6] Pour une histoire du développement de la génétique en France dans l'entre-deux guerres, on pourra se reporter à (Gayon et Burian, 2004). Pour une comparaison des rapports entre mendélisme et médecine en Allemagne, en Angleterre et en France, on pourra se reporter à (Gaudillière, 2000). Enfin, pour une histoire de la construction de la génétique médicale comme discipline institutionnelle dans les années 1970, se référer à (Coventry et Pickstone, 1999)

[7] Pour des exemples, on se reportera aux analyses de Charles Scriver (Scriver, 1995, 2007 ; Scriver et Waters, 1999) et de Diane Paul (Paul, 1994, 2000, 2013) sur la phénylcétonurie ou aux analyses d'Ann Kerr (Kerr, 2000, 2004, 2005) , Adam Hedgecoe (Hedgecoe, 2003) ou Lindee (Lindee et Mueller, 2011) sur l'anémie falciforme.

D'abord parce que le porteur d'une maladie infectieuse (sauf dans le cas éventuel d'une femme enceinte) n'a pas à se préoccuper de sa maladie dans l'élaboration de son projet parental, dans la déclaration à l'assurance ou dans le cadre d'un entretien professionnel, quoique ces deux derniers problèmes ne soient pas d'actualité en France. Mais aussi parce qu'avec le battage médiatico-politique qui a lieu autour de la génétique et des maladies génétiques, être atteint d'une maladie génétique a des connotations sociales tout à fait particulières (sentiment d'un « défaut intime », impression de destin… (Dekeuwer, 2006)) . Enfin, très peu de maladies génétiques ont des traitements curatifs, même si certains traitements palliatifs ou de confort peuvent soulager les symptômes de la maladie : le diagnostic de maladie génétique résonne souvent comme une condamnation. La deuxième considération de Bowker et Star est que les classifications sont au moins en partie le reflet d'un concept. Ainsi, la façon dont sont classées les maladies génétiques est intrinsèquement reliée, comme on va le voir dans la suite de cet exposé, à des considérations sur la nature du matériel génétique, son rôle étiologique dans la détermination de la maladie et des autres facteurs susceptibles d'intervenir. Les évolutions d'une classification sont ainsi le reflet de l'évolution d'un concept. Pour ne donner qu'un exemple, si dans les années trente, un patient atteint de la chorée de Huntington se serait trouvé catalogué comme « dément » à l'Institut de Colchester à côté des autistes, des schizophrènes, des alcooliques, des hystériques, la chorée de Huntington est aujourd'hui considérée comme une maladie neurodégénérative, ce qui a un impact dans sa prise en charge diagnostique et thérapeutique. Dernière considération enfin : toute classification est orientée par d'autres critères que les critères ontologiques. Si les classifications scientifiques sont supposées révéler la véritable nature des objets qu'elle classe, il est bien évident que ces classifications n'en sont pas moins soumises à un ensemble de contraintes extérieures à la nature de leur objet (contexte de la construction de la classification, public visé par la classification, présupposés sociaux et culturels) qui participent aussi à leur édification et à leur évolution. Par exemple, reclasser une maladie comme génétique peut s'intégrer à une stratégie de recherche de subventions pour un laboratoire scientifique. Pour donner un exemple moins controversé, la classification des maladies génétiques qui est présentée sur le site du Projet Génome Humain dans un but de vulgarisation n'a pas le même public que les classifications

proposées dans des manuels de génétique médicale. C'est pourquoi nous serons attentifs dans la présentation des classifications que nous étudierons au contexte de leur élaboration ainsi qu'à leurs objectifs.

1.1.2. Quelle place pour les maladies génétiques au sein de la Classification Internationale des Maladies (CIM-10)?

Il n'existe pas de classification officielle des maladies génétiques. Par « officielle », nous voulons dire qu'il n'existe pas de consensus scientifique décidé par les instances de santé publique sur une classification des maladies génétiques. Dans ce contexte, notre premier réflexe a été de nous tourner vers la Classification Internationale des Maladies : peut-être y avait-il à l'intérieur de cette classification une sous-classification des maladies génétiques qui rendrait inutile et superfétatoire toute classification indépendante des maladies génétiques ? Parce que tel n'est justement pas le cas et que les maladies génétiques au sein de la Classification Internationale des Maladies font l'objet d'une étonnante dispersion et non d'une catégorie à part entière, il nous a paru intéressant de faire un bref détour par une analyse de la CIM-10.

1.1.2.1. Une brève histoire de la Classification Internationale des Maladies

L'histoire de la Classification statistique internationale des maladies et des problèmes de santé connexes (dont l'abréviation courante est CIM) a commencé avec la classification des causes de décès de Jacques Bertillon en 1938. Après cinq révisions décennales (la cinquième date de 1938), cette classification qui répertoriait uniquement les causes de mortalité a connu un changement majeur en s'intéressant à partir de sa sixième révision en 1946 aux causes de morbidité et donc à la classification des maladies. Ce travail de codage et de classification, dont nous expliciterons les principes un peu plus loin est alors confié à l'Organisation Mondiale de la Santé (OMS) et elle devient la « Classification statistique internationale des maladies, traumatismes et causes de décès ». L'OMS a brisé le rythme des révisions décennales avec la dernière version de la CIM : étant donné l'importance des modifications à apporter, la CIM-10 n'a vu le jour qu'en 1993 et fait l'objet depuis de constantes mises à jour en ligne. L'arrivée officielle de la CIM-11 est prévue pour 2015. Une première version dite

« alpha » a été créée par un groupe restreint de centres de recherche. Depuis 2012, une version ouverte au public mais qui n'est pas encore officielle, dite « beta », est accessible en ligne et fait l'objet de contributions du public. C'est parce que la CIM-11 est encore en cours de révision que nous nous sommes concentrée sur l'analyse de la version précédente.

1.1.2.2. Présentation de la CIM -10

La CIM est composée de trois volumes parus respectivement en 1993, 1995 et 1996. Le volume 1 est la Table analytique, autrement appelée « classification systématique », qui se découpe en vingt-deux chapitres (depuis la dernière mise à jour en 2006). Les maladies sont classées en fonction d'un code alphanumérique (c'est-à-dire un code composé de toutes les lettres de l'alphabet et d'une suite de chiffre de 0 à 9). Le volume 2 est le Manuel d'utilisation, qui explicite l'histoire et les règles de la CIM et qui détaille la façon dont elle peut être utilisée. Le volume 3 est un index alphabétique : on y trouve les maladies répertoriées dans le volume 1 mais le critère est non pas le codage de la maladie ou de la cause de décès mais le nom de la maladie.

Chacun des vingt-deux chapitres composant la table analytique comporte des blocs qui regroupent un ensemble de catégories homogènes. Ces catégories sont désignées par un code à trois caractères, une lettre et deux chiffres. Une catégorie est créée pour des situations simples ou des groupes de maladies avec des caractéristiques communes. De plus, il est prévu une place pour "d'autres" situations relatives à la même catégorie, ainsi que pour des situations "non spécifiées". Le titre du bloc explicite l'homogénéité des catégories regroupées. Ou bien un bloc représente un ensemble de catégories, auquel cas son étendue est déterminée par la catégorie de début et la catégorie de fin du groupement ; ou bien un bloc représente lui-même une catégorie et dans ce cas-là, le code indiquant la catégorie est indiqué entre parenthèses. Par exemple, le premier chapitre de la CIM-10 est : « Certaines maladies infectieuses et parasitaires » et il comprend les catégories A00 à B99. Il comprend par ailleurs vingt et un blocs, dont le premier est A00-A09 et dont le titre est « Maladies infectieuses parasitaires ». Dans ces vingt-et-un blocs, on trouve aussi des blocs qui

représentent une seule catégorie comme le groupe (B99), dénommé « Autres maladies infectieuses » :

Chap I : A00-B99 : Certaines maladies infectieuses et parasitaires

A00-A09 : Maladies infectieuses parasitaires

….

(B99) : Autres maladies infectieuses

Chacune des catégories peut alors se subdiviser en sous-catégories dont le code comprend un quatrième caractère séparé des trois premiers par un point.

1.1.2.3. Une classification statistique et pragmatique, donc incomplète et ambiguë

Deux caractéristiques de la Classification Internationale des Maladies méritent d'être soulignées. D'une part, la CIM ne cherche pas à répertorier toutes les maladies existantes, elle ne répertorie que les maladies dont la fréquence et l'extension géographique en font des problèmes de santé publique. C'est donc une classification statistique, qui n'a pas pour objectif l'exhaustivité. La précision avec laquelle les catégories sont détaillées et les états et maladies répertoriés dépend de leur gravité et de leur fréquence à l'état mondial : le chapitre sur les maladies infectieuses et parasitaires est ainsi particulièrement détaillé parce que ces maladies sont considérées comme une des premières causes de mortalité dans le monde. D'ailleurs, il faut préciser que la classification internationale des maladies est complétée par un certain nombre de classifications supplémentaires et partiellement redondantes : des classifications par spécialité, la classification internationale du fonctionnement, du handicap et de la santé (CIF), la classification internationale des interventions en santé et enfin, la nomenclature internationale des maladies dont le but est d'attribuer un nom et un seul à chaque maladie.

Cette non-exhaustivité tient en partie au caractère essentiellement pragmatique de la Classification Internationale des Maladies, qui se conçoit avant tout comme un outil d'aide pour les politiques publiques de la santé et l'épidémiologie :

> « La CIM s'est développée comme une classification pratique plutôt que comme une classification purement théorique, dans laquelle il y a un certain

> nombre de compromis entre une classification fondée sur l'étiologie, la localisation anatomique, les circonstances d'apparition, etc… Des ajustements ont été également nécessaires pour répondre à la variété des applications statistiques pour laquelle l'ICD est réalisée, comme la mortalité, la morbidité, la sécurité sociale et toutes les autres statistiques et enquêtes qui concernent la santé. » (OMS, 2006, p. 10)

Deux aspects nous intéressent dans cette citation. D'une part, nous voudrions souligner que l'importance accordée à la santé publique, si elle se comprend parfaitement dans la perspective d'une classification internationale des maladies, peut du même coup être une source de conflits à l'échelle individuelle. D'autre part, la question des compromis entre principes de classification nous paraît particulièrement intéressante. Dans leur livre *Sorting things out : classification and its consequences,* Bowker et Star se sont tout particulièrement intéressés à la question de la classification internationale des maladies, et ils considèrent que celle-ci représente effectivement un compromis entre plusieurs principes de classification, d'où une ambiguïté sur le statut de la CIM : est-ce une classification ou une nomenclature ? En effet, on distingue généralement la classification de la nomenclature sur la question des principes de classement. Une classification est fondée sur un seul principe de classification qui doit être cohérent, dont les catégories sont mutuellement exclusives et qui aboutit à un système complet. Une nomenclature est simplement une liste de noms sans principes de classification et sur lesquels il y a consensus. La question de savoir si la CIM est une classification ou une nomenclature est une question récurrente depuis sa création. Dans le manuel d'utilisation de la CIM, on trouve ainsi cette formule extrêmement ambiguë :

> « C'est le fait de regrouper des catégories qui distingue la classification statistique d'une nomenclature, dans laquelle chaque état morbide connu doit avoir un titre distinct. Les concepts de classification et de nomenclature sont cependant étroitement liés dans la mesure où une nomenclature est souvent arrangée de façon systématique. » (OMS, 2006, p. 4)

Formellement, la CIM, tout en se considérant comme une classification, est pourtant constituée de plusieurs principes de classification. Ces principes de

classification non mutuellement exclusifs (d'où le recours au système d'inclusions et exclusions et le système d'appariement des codes) étaient historiquement inévitables dans la mesure où les usages de la CIM sont multiples et où certains groupes revendiquent en fonction de leur intérêt propre des principes de classification différents : *« La CIM s'est construite de façon aussi hétérogène que possible pour permettre aux différents groupes d'y refléter leurs propres soucis. Parce que différents modèles médicaux se maintiennent, différentes règles de classement sont incorporées.»* (Bowker et Star, 2000, p. 150 - 151). Cette ambiguïté est palpable dans la liste des codes de la CIM-10 (Tableau 1) :

I A00-B99 Certaines maladies infectieuses et parasitaires
II C00-D48 Tumeurs
III D50-D89 Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire
IV E00-E90 Maladies endocriniennes, nutritionnelles et métaboliques
V F00-F99 Troubles mentaux et du comportement
VI G00-G99 Maladies du système nerveux
VII H00-H59 Maladies de l'œil et de ses annexes
VIII H60-H95 Maladies de l'oreille et de l'apophyse mastoïde
IX I00-I99 Maladies de l'appareil circulatoire
X J00-J99 Maladies de l'appareil respiratoire
XI K00-K93 Maladies de l'appareil digestif
XII L00-L99 Maladies de la peau et du tissu cellulaire sous-cutané
XIII M00-M99 Maladies du système ostéo-articulaire, des muscles et du tissu conjonctif
XIV N00-N99 Maladies de l'appareil génito-urinaire
XV O00-O99 Grossesse, accouchement et puerpéralité
XVI P00-P96 Certaines affections dont l'origine se situe dans la période périnatale
XVII Q00-Q99 Malformations congénitales et anomalies chromosomiques
XVIII R00-R99 Symptômes, signes et résultats anormaux d'examens cliniques et de laboratoire, non classés ailleurs
XIX S00-T98 Lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes
XX V01-Y98 Causes externes de morbidité et de mortalité
XXI Z00-Z99 Facteurs influant sur l'état de santé et motifs de recours aux services de santé
XXII U00-U99 Codes d'utilisation particulière

Tableau 1 : Liste des principaux codes de la CIM-10

Le premier chapitre « Certaines maladies infectieuses et parasitaires » fait référence à un principe étiologique : il s'agit d'identifier la cause de la maladie en faisant référence au modèle du germe qui a émergé à l'ère pasteurienne. Les chapitres III, IV, VI à XIV font indiscutablement référence à un principe topographique : les

maladies sont classées par leur localisation dans le corps humain, en désignant l'organe ou le système touché. Les chapitres XV, XVI et XVII sont ambigus : pris ensemble, ils semblent désigner un moment clinique : il s'agit de répertorier toutes les maladies qui sont rattachées d'une façon ou d'une autre à la grossesse, qu'elles concernent le moment de la grossesse, le moment de l'accouchement ou le moment de la naissance, que le patient soit la mère, l'embryon, le fétus ou l'enfant. Pour les chapitres XV et XVI, on peut quasiment parler de classement par spécialité dans la mesure où les maladies du XV concernent uniquement la mère et la lettre O désigne l'obstétrique, le chapitre XVI concerne uniquement l'enfant et la lettre P désigne la pédiatrie. On retrouve cependant la pédiatrie dans le chapitre XVII qui touche aussi la génétique médicale. On ne peut pas parler de classement étiologique dans la mesure où les maladies décrites dans le chapitre XV par exemple font référence aussi bien à l'avortement qu'à des infections virales susceptibles de mettre en danger la mère. Nous serions tentés de nommer ce principe « contextuel », car c'est peut-être le « moment clinique » ici qui fait l'unité de classement. On pourrait aussi parler de principe opérationnel : clairement, les trois chapitres font référence à des maladies qui sont souvent traitées dans un regroupement hospitalier associant un service d'obstétrique, une maternité et un service de génétique médicale.

1.1.2.4. Les maladies génétiques dans la CIM-10, une dispersion étonnante

Premier constat, les maladies génétiques ne forment pas une catégorie étiologique distincte, comme c'est le cas pour « certaines maladies infectieuses et parasitaires ». Un chapitre semble néanmoins concerner au moins une partie d'entre elles, c'est le chapitre XVII « Malformations congénitales et anomalies chromosomiques » (Tableau 2). Le chapitre ne contient qu'une seule lettre, ce qui suppose que la catégorie « maladies congénitales et anomalies chromosomiques » est bien comprise comme un continuum et non comme le regroupement de deux catégories distinctes dans un chapitre. L'onglet désigne donc toutes les maladies ou états médicaux qui sont présents à la naissance, qu'ils aient une cause génétique, soient dus à l'environnement intra-utérin ou à des facteurs extérieurs. Dans ce chapitre, on trouve ainsi aussi bien : « Q 86.0 Syndrome d'alcoolisme fœtal » dont la

cause est exogène (consommation d'alcool par la mère), « Q 90 Syndrome de Down », une anomalie chromosomique sporadique (non transmise par les parents, mais apparaissant *de novo*) et qui se définit par la présence partielle ou totale d'un troisième chromosome 21 et enfin, « Q 35-Q37 : les fentes labiales et palatines » dont certaines peuvent avoir une origine héréditaire (par exemple, le risque d'avoir un enfant affligé d'un bec de lièvre est augmenté par 35 si l'un des parents en présente un). Les maladies congénitales sont donc classées non étiologiquement mais topographiquement. Où trouve-t-on d'autres maladies habituellement considérées comme des « maladies génétiques » ?

Q00-Q07 Malformations congénitales du système nerveux
Q10-Q18 Malformations congénitales de l'œil, de l'oreille, de la face et du cou
Q20-Q28 Malformations congénitales de l'appareil circulatoire
Q30-Q34 Malformations congénitales de l'appareil respiratoire
Q35-Q37 Fente labiale et fente palatine
Q38-Q45 Autres malformations congénitales de l'appareil digestif
Q50-Q56 Malformations congénitales des organes génitaux
Q65-Q79 Malformations congénitales du système ostéo-articulaire et des muscles
Q60-Q64 Malformations congénitales de l'appareil urinaire
Q80-Q89 Autres malformations congénitales
Q90-Q99 Anomalies chromosomiques, non classées ailleurs

Tableau 2 : Liste des principaux codes du chapitre XVII « Malformations congénitales et anomalies chromosomiques » de la CIM-10

Le chapitre « Maladies congénitales et anomalies chromosomiques » exclut les « erreurs innées et les anomalies du métabolisme ». Ce bloc E70-E90 fait partie du chapitre sur les « Maladies endocriniennes, nutritionnelles et métaboliques » (chapitre IV) et comprend effectivement des maladies classiquement considérées comme des maladies génétiques. On y trouve ainsi une des « maladies-stars » de la génétique médicale, la phénylcétonurie (E70). En fait, on réalise que « le reste » des maladies génétiques est classé topographiquement en fonction de l'organe ou du système touché. Ceci est valable aussi bien pour des maladies stars de la génétique (la drépanocytose est codée D57, dans le chapitre III « Maladies du sang et des organes

hématopoïétiques et certaines maladies du système immunitaire ») ou pour des maladies dont le caractère génétique est davantage controversé (la schizophrénie par exemple est codée F20.9 dans le chapitre V « Troubles mentaux ou du comportement »). Un exemple intéressant est la codification de la forme familiale du cancer du sein : il fait en effet l'objet d'une double classification. Classé en C50.9, il est alors désigné comme une « Tumeur maligne du sein » dans le chapitre II « Tumeurs ». Classé en Z80.3, dans le chapitre XXI « Facteurs influant sur l'état de santé et motifs de recours aux services de santé », il apparaît dans le bloc « Sujets dont la santé peut être menacée en raisons d'antécédents personnels et familiaux et de certaines affections ». Le classement de la forme familiale du sein est donc à la fois topographique et opérationnel. Par opérationnel, nous voulons dire que la catégorie « Sujets dont la santé peut être menacée en raison d'antécédents personnels et familiaux de certaines affections » désigne des sujets à risque nécessitant une surveillance médicale et clinique (dépistage, mammographie,…) plus importante que les autres. Il est intéressant aussi de noter que cette catégorie semble regrouper les maladies familiales, c'est-à-dire les maladies dont la transmission est héréditaire, même si le type d'hérédité ou le mode de transmission de ces maladies n'a pas été identifié.

1.1.2.5. Les raisons de la dispersion

Que peut-on dire de la classification des maladies génétiques dans la CIM ? Essentiellement que les maladies génétiques ne font pas l'objet d'une classification propre mais sont dispersées dans plusieurs blocs et chapitres : dans le chapitre sur les maladies congénitales et les anomalies chromosomiques, dans le chapitre portant sur les maladies familiales, et un peu partout dans chaque chapitre en fonction de la localisation topographique. Il n'y a donc pas de principe unique de classification des maladies génétiques : les maladies génétiques ne constituent pas dans la CIM une catégorie étiologique, ni une catégorie topographique, ni même une catégorie opérationnelle.

Pourquoi cette dispersion ? Au moins deux explications peuvent être proposées. D'une part, la CIM est une classification qui a fait l'objet de nombreuses révisions en fonction de l'évolution des connaissances scientifiques mais une

contrainte forte s'exerce sur ces révisions : chaque changement et chaque révision impliquent des difficultés de correspondance et de comparaison statistique entre la nouvelle version de la CIM et l'ancienne. Ce problème s'est posé avec beaucoup d'acuité entre la CIM-9 et la CIM-10 : étant donné l'ampleur des modifications entre les deux versions, les comparaisons statistiques dans une perspective épidémiologique ou de politiques publiques sont très délicates. Quand on connaît cet élément et qu'on sait que la génétique médicale a connu son plus fort développement après la Seconde Guerre mondiale, on mesure à quel point il peut être complexe de créer un chapitre spécifique pour les maladies génétiques et à quel point la CIM s'en trouverait bouleversée.

Une autre de nos hypothèses que nous aurons l'occasion de détailler un peu plus loin est qu'il est très difficile de parvenir à une définition simple de ce qu'est une maladie génétique. C'est d'ailleurs ce que reflète la dispersion des maladies génétiques dans la CIM : la notion de maladie génétique recoupe partiellement celles de maladie familiale ou héréditaire, mais elle comprend aussi les anomalies chromosomiques qui sont très rarement héritées des parents et sont souvent l'objet d'aberrations sporadiques dans la structure ou le nombre des chromosomes. Enfin, il existe un grand nombre de maladies dont l'étiologie est mal connue ou sujette à discussion : par exemple, le *spina bifida* est une malformation congénitale liée à un défaut de fermeture du tube neural durant la vie embryonnaire, qui reste ouvert à son extrémité caudale. Son origine reste sujette à controverse, d'aucuns penchent pour une origine héréditaire, d'autres pour une déficience d'acide folique. Les exemples les plus frappants et les plus controversés en termes de nosologie médicale sont certaines maladies mentales comme l'autisme ou la schizophrénie où la part revenant aux gènes et celle revenant à l'environnement, à la société et à des causes exogènes, sont l'objet de débats perpétuels. Le point important à ce stade de notre propos est que si la classification internationale des maladies ne contient pas de classification homogène officielle des maladies génétiques, des classifications officieuses spécifiques ont fleuri. Nous allons à présent nous intéresser à une de ces classifications officieuses : la base de données *Online Inheritance on Man*.

1.2. Les maladies génétiques dans la base de données *Online Mendelian Inheritance on Man* (OMIM)

1.2.1. OMIM, classification « officieuse » des maladies génétiques

1.2.1.1. Victor McKusick, fondateur d'OMIM

Victor McKusick (1921-2008), surnommé le « père de la génétique médicale » (Collins, 2008 ; Valle, 2008), a contribué au développement et à la diffusion de cette discipline aussi bien au niveau institutionnel et universitaire qu'au niveau de la recherche biomédicale, alors même que la scientificité et l'intérêt de celle-ci était loin de faire l'unanimité (Harper, 2012) .

Sur le plan institutionnel et universitaire, McKusick a créé en 1953 une Chaire de Génétique Médicale à l'Institut Johns Hopkins, qui est aujourd'hui un des centres de génétique médicale les plus réputés dans le monde. Pour ne citer que quelques-unes des distinctions les plus récentes qui ont récompensé son travail au cours de sa longue carrière, on peut parler du prix Albert Lasker qui salue les réussites scientifiques les plus spectaculaires dans les sciences médicales en 1997, mais aussi de la Médaille Nationale de Science en 2001, du Prix Japonais de Génomique Médicale et de Génétique en 2008. Il a été le président fondateur de l'Organisation pour le Génome Humain (HuGO) et membre de l'Académie américaine Nationale des Sciences. Il a également dirigé deux journaux spécialisés (*Medicine* et *Genomics*). Ceci dit, on considère souvent que sa plus belle réussite, celle qui a aidé au mieux au développement de la recherche biomédicale en génétique et en génomique est la constitution de la base de données OMIM.

1.2.1.2. La naissance de MIM

Médecin spécialisé en cardiologie, McKusick décrit lui-même son intérêt pour la nosologie et la nosographie des maladies génétiques à travers trois grands travaux qui ont fait sa renommée : le catalogue des désordres héréditaires des tissus conjonctifs en 1956, la cartographie du chromosome X et ses travaux sur les maladies autosomiques récessives chez les Amish. Ce sont ces trois approches qui le conduisent

à publier la première édition du catalogue « Hérédité mendélienne chez l'homme. Catalogue des phénotypes autosomiques dominants, autosomiques récessifs et X-liés » en 1966, comme Victor McKusick le rappelle dans un article de 2007 publié dans *The American Journal of Genetics* et où il revient sur l'origine de ce catalogue quarante années auparavant :

> « MIM est né de trois tentatives différentes ; en premier lieu, la publication annuelle de la revue de génétique médicale que moi et mes collègues à la Clinique Moore de l'institut Johns Hopkins avons préparée durant six années de 1958 à 1963 ; en second lieu, le catalogue des traits X-liés que j'ai compilés pour estimer le contenu génétique du chromosome X ; en troisième lieu, le catalogue des phénotypes autosomiques récessifs que j'ai utilisé en 1963 comme ressource pour identifier à la fois les « anciennes » et les « nouvelles » maladies récessives dans mes études de l'ancienne tribu des Amish. » (McKusick, 2007, p. 590)

Dans un premier temps, le catalogue est publié et édité par le *Johns Hopkins Institute*. Le catalogue s'inscrit dans un triple objectif – classer les causes de la maladie, décrire les entités cliniques que sont les maladies et attribuer à chaque maladie un nom univoque :

> « La nosologie (qui signifie littéralement « l'étude des maladies » mais qui est traditionnellement compris comme la classification et la délimitation des maladies) et la nosographie (« la description de la maladie ») et la nomenclature des maladies ont été nécessairement des considérations centrales tout au long de l'histoire de MIM. En effet, la nosologie et la nosographie que je comprends comme le fait de délimiter des traits génétiques distincts, ont formé les principales fondations de l'assemblage de MIM. » (McKusick, 2007, p. 591–592)

En effet, le but de Victor McKusick était d'offrir une liste de ressources détaillées sur les connaissances toujours changeantes de la génétique médicale à un public de spécialistes, de médecins, d'étudiants et de chercheurs. C'est cet objectif qui explique en grande partie l'importance des révisions effectuées pendant les quarante dernières années sur MIM. Ces révisions étaient néanmoins longues, relativement

compliquées sur le plan informatique et très importantes du fait de l'explosion des techniques de la biologie moléculaire à partir des années soixante-dix.

1.2.1.3. De MIM à OMIM, du catalogue à la base de données

Un tournant décisif a eu lieu dans les années quatre-vingt, lorsque la Bibliothèque Nationale de Médecine a développé à la fois un logiciel d'extraction des données et un moteur de recherche qui ont permis la mise en ligne de la sixième édition du catalogue (1983) dès l'année 1987. Cette transformation a eu plusieurs conséquences de premier plan :

- ***Une simplification des processus de révision :*** aujourd'hui les mises à jour sont quotidiennes et gérées en grande partie par les scientifiques du *Johns Hopkins Institute*. Étant donné l'augmentation du nombre d'entrées ces dernières années, on comprend bien l'importance d'une mise à jour quotidienne des données contenues dans OMIM (Figure 1) :

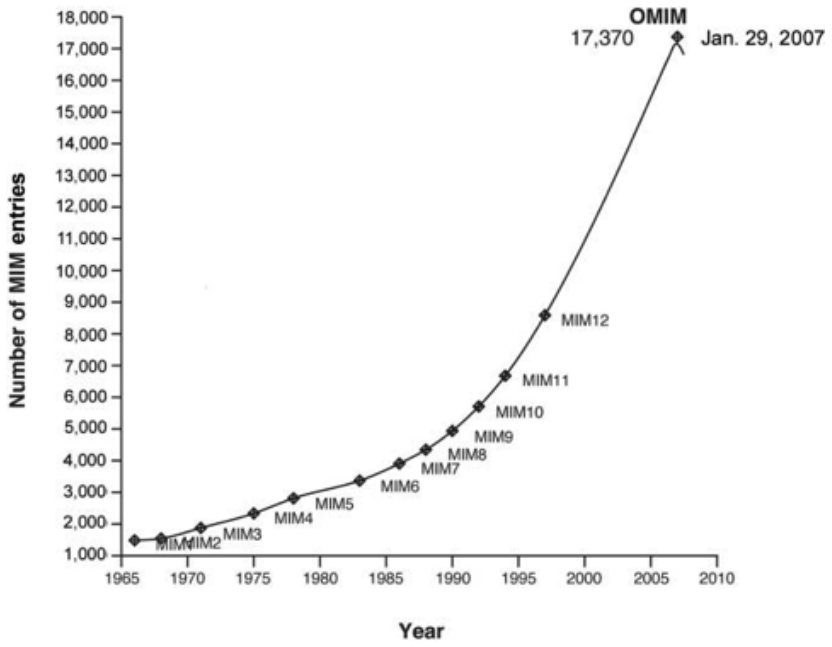


Figure 1 : Évolution du nombre d'entrées dans OMIM depuis sa création (McKusick, 2007) .

- ***Une mise en réseau et un partage des ressources sur la génétique médicale :*** Une des grandes forces d'OMIM est effectivement son système de références croisées

(liens hypertextes entre les différentes entrées), de références externes (liens vers des articles de référence sur la base de données PubMed, liens vers la nomenclature HUGO, liens vers GeneTests qui répertorie les différents types de tests génétiques actuellement disponibles dans le monde et les centres dans lesquels sont disponibles ces tests, liens vers les bases de données donnant la séquence de l'ADN et de la protéine impliqués, liens vers les bases de données sur les mutations générales et spécifiques à un locus, liens vers plusieurs associations de patients, liens vers GeneMap qui permet de visualiser la localisation topographique d'un gène impliqué dans un phénotype pathogène sur le ou les chromosomes impliqués, etc.).

- ***Une augmentation de la fréquentation :*** Elle a permis de toucher un public beaucoup plus important avec une navigation beaucoup plus simple et rapide que sur le catalogue en version imprimée. Quelques chiffres peuvent donner une idée de l'importance du phénomène : en 2005, on compte en moyenne 8 500 utilisateurs individuels pour 100 000 recherches par jour (Hamosh *et al*., 2005).

- ***Une meilleure reconnaissance scientifique*** : Le fait que l'édition en ligne du catalogue ait été successivement prise en charge par la *National Library of Medicine* (NLM) depuis 1987 et par le *National Center for Biotechnology Information* depuis 1995 (NCBI), deux institutions extrêmement célèbres et reconnues des États-Unis, a donné une légitimité scientifique à cette base de données. OMIM est ce qui se rapproche le plus d'une classification « officielle » des maladies génétiques, reconnue internationalement comme l'est la Classification Internationale des Maladies publiée par l'OMS par exemple.

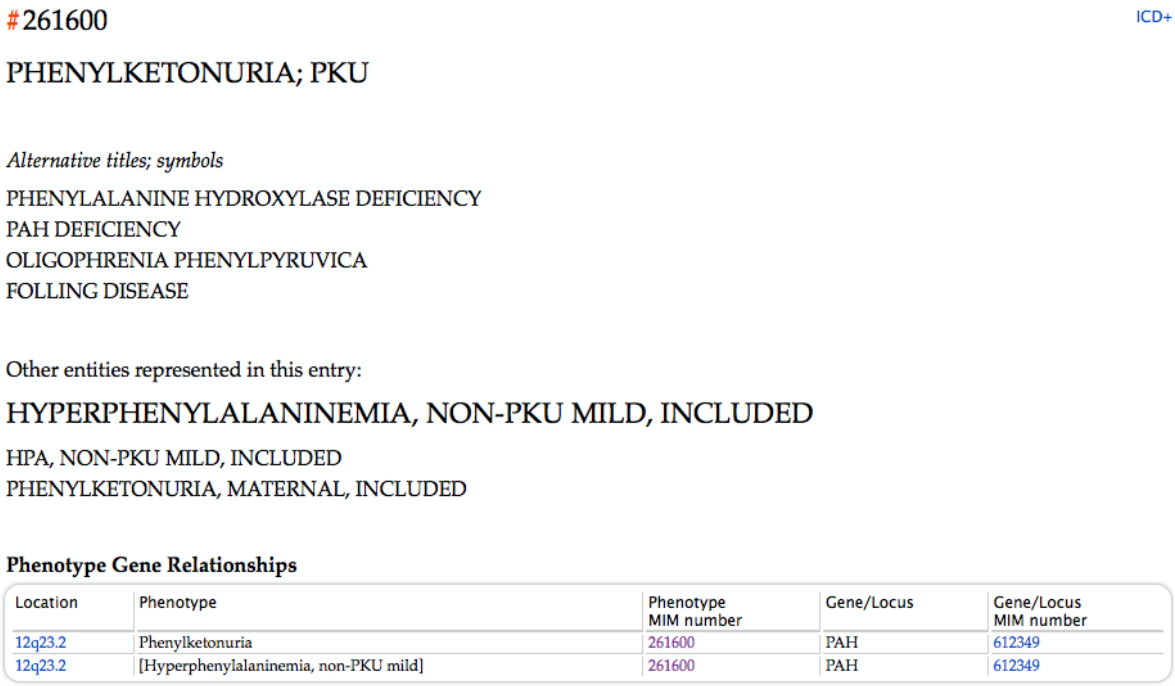
Il nous a semblé particulièrement important de nous attarder sur les changements représentés par le passage du catalogue papier à la base de données en ligne, non seulement parce que ces transformations rendent compte du statut particulier d'OMIM, devenue une véritable « classification officieuse » des maladies génétiques, mais aussi pour justifier le fait que nous préférons étudier la classification de Victor McKusick sous sa forme en ligne plutôt qu'en version papier. En effet, la dernière version imprimée à ce jour est la douzième édition du catalogue en 1998. Depuis, plus aucune version papier n'a été publiée, étant donné les extraordinaires

améliorations que permet la version en ligne et on peut gager que cette douzième édition sera bien la dernière version imprimée du catalogue de McKusick.

1.2.2. Description d'une entrée dans OMIM

L'objectif initial de Victor McKusick était de proposer un ensemble de corrélations entre génotype et phénotype. Comme le sous-entend donc le nouveau sous-titre que le catalogue a pris depuis l'édition de 1994 (« Un catalogue des gènes humains et des désordres génétiques »), OMIM recense au moins deux types d'objets : d'une part, des phénotypes la plupart du temps pathogènes et d'autre part, des gènes qui peuvent être désignés par leurs loci ou déclinés en fonction de leurs variants alléliques.

Pour chaque entrée, on trouve une description dont le niveau de détail dépend du niveau de connaissances scientifiques du moment, mais qui prend en compte lorsque cela est possible l'histoire de la découverte du locus ou du phénotype, l'épidémiologie du locus ou du phénotype (fréquence, population spécifiquement touchée,…) la description moléculaire du polymorphisme repéré, la description du phénotype associé, des renvois vers des articles de références publiés sur PubMed, les indications thérapeutiques ou les essais cliniques à la recherche d'une thérapie en cours. Ce sont donc des articles de « plusieurs pages » qui sont régulièrement mis à jour pour constituer une excellente synthèse des connaissances scientifiques contemporaines sur un phénotype ou un gène donné. Voici un exemple de l'entrée qui concerne la phénylcétonurie (Figure 2). Le nom de la maladie, ainsi que les autres noms usuels sont indiqués en dessous du numéro à six chiffres correspondant (ici MIM #261600). Dans le tableau sont indiquées les relations existant entre phénotype et génotype avec la localisation du gène et le numéro correspondant à l'entrée du gène.



#261600 ICD+

PHENYLKETONURIA; PKU

Alternative titles; symbols

PHENYLALANINE HYDROXYLASE DEFICIENCY
PAH DEFICIENCY
OLIGOPHRENIA PHENYLPYRUVICA
FOLLING DISEASE

Other entities represented in this entry:

HYPERPHENYLALANINEMIA, NON-PKU MILD, INCLUDED

HPA, NON-PKU MILD, INCLUDED
PHENYLKETONURIA, MATERNAL, INCLUDED

Phenotype Gene Relationships

Location	Phenotype	Phenotype MIM number	Gene/Locus	Gene/Locus MIM number
12q23.2	Phenylketonuria	261600	PAH	612349
12q23.2	[Hyperphenylalaninemia, non-PKU mild]	261600	PAH	612349

Figure 2 : Capture d'écran de la page "titre" de l'entrée "phénylcétonurie" sur le site OMIM. Source : http://omim.org/entry/261600. (Capture d'écran du 7 février 2013).

Par ailleurs, une table des matières de l'entrée (Figure 3) indique l'ensemble des informations disponibles pour cette entrée codée « MIM #261600 ». En l'occurrence, pour la phénylcétonurie, sont disponibles la description de la maladie, la liste des symptômes cliniques et biochimiques, le mode d'hérédité, la carte génétique, une description en termes de génétique moléculaire, les corrélations génotype/phénotype, la pathogénèse, le diagnostic, la prise en charge clinique, l'épidémiologie et l'histoire évolutionnaire de la maladie. Enfin, en dessous de la table des matières, on trouve un onglet (reproduit ci-dessous – Figure 3) qui renvoie vers des liens externes permettant d'accéder à des ressources cliniques (telles que des bases de données des mutations du gène impliqué dans la phénylcétonurie, des bases de données sur les tests génétiques ou sur les politiques de dépistage), mais aussi des liens vers les modèles animaux ou les phénomènes cellulaires et métaboliques impliqués dans la maladie.

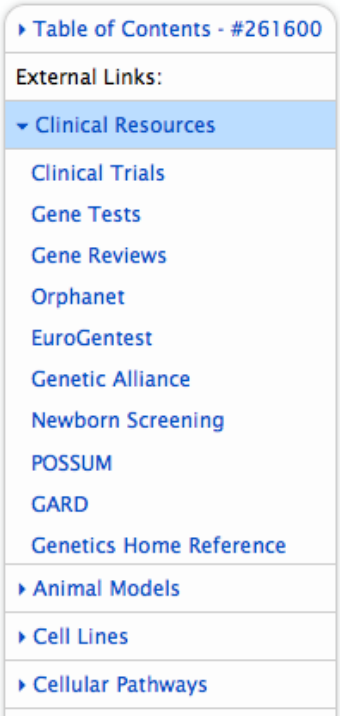


Figure 3 : Capture d'écran de l'onglet renvoyant aux liens externes accessibles à partir d'OMIM pour la phénylcétonurie. Source : http://omim.org/entry/261600. (Capture d'écran du 7 février 2013).

Pour terminer cette présentation d'OMIM, il nous faut présenter le double système de classement utilisé pour organiser les plus de vingt-et-un mille entrées qu'elle contient à ce jour : d'une part, le système de numérotation qui attribue un numéro à tout phénotype et à tout gène indiquant son mode de transmission et d'autre part un système de symboles qui permet d'identifier la nature de l'objet référencé. C'est l'évolution de ce double système de classement depuis 1966 qu'il est particulièrement intéressant de retracer pour comprendre comment le concept de maladie génétique a évolué pendant cette période.

1.2.3. Le modèle de la maladie génétique aux débuts d'OMIM

Avant de décrire les transformations qu'a subies OMIM, il nous a paru opportun de présenter ce qui est considéré comme le modèle de la maladie génétique aux débuts d'OMIM – précisément afin de mesurer l'étendue des transformations de ce modèle. Le catalogue de McKusick ayant été édité en 1966, nous avons choisi de nous concentrer sur la phénylcétonurie, maladie-star de la génétique médicale dans les années soixante, que nous avons déjà brièvement évoquée en introduction et qui est l'une des premières maladies génétiques pour laquelle ont été proposés un dépistage

néonatal et un traitement dès septembre 1965 aux États-Unis. En 1934, en examinant deux enfants pour rechercher la cause de leur retard mental, le norvégien Asbjørn Følling met en évidence la présence d'un métabolite particulier dans leur urine, l'acide phénylpyruvique. Ce désordre inné du métabolisme soulève l'intérêt de Lionel Penrose, qui en fait le sujet de son allocution inaugurale à l'Université Galton en Angleterre en 1946 et pose alors de nombreuses hypothèses : la phénylcétonurie est décrite comme un désordre autosomique récessif du métabolisme de la phénylalanine concernant environ un enfant sur quinze mille naissances et causant un retard mental (Penrose, 1998). Ce n'est qu'en 1953 que la maladie est identifiée formellement comme une déficience de l'enzyme hépatique, la phénylalanine hydroxylase. L'explication de la phénylcétonurie est alors la suivante : une mutation dans le gène responsable de la production de l'enzyme phénylalanine hydroxylase aboutit à la non production de l'enzyme PAH qui convertit la phénylalanine (un acide aminé qu'on trouve dans l'alimentation) en tyrosine. Comme il n'y a plus d'enzyme de conversion de la phénylalanine en tyrosine, la phénylalanine et ses métabolites s'accumulent et passent la barrière hémato-encéphalique, entraînant un retard cognitif et mental. Entre 1960 et 1963, est alors mis au point un test de dépistage particulièrement simple (le test de Guthrie qui permet de mesurer le taux de phénylalanine plasmatique sur une goutte de sang). Les premières tentatives de traitement par régime sans phénylalanine (dont Lionel Penrose avait déjà fait l'hypothèse en 1946) sont des succès et vont donner lieu très rapidement à une politique de dépistage néonatal. Comme le résume Susan Lindee :

> « En une dizaine d'années, de 1953 à 1963, un obscur désordre métabolique a émergé de ce que Edward Yoxen a appelé le « ghetto médical » des maladies génétiques pour devenir le centre biochimique d'un vaste réseau d'information, de technologie, de prise en charge sociale et de législation. Un rare désordre du métabolisme de la phénylalanine – la phénylcétonurie (PKU, OMIM 261 600) – est devenu le centre d'une importante initiative de santé publique. Concernant environ une naissance sur quinze mille, il a attiré l'attention et les financements d'organisations consacrées au contrôle du retard mental chez les enfants. Des douzaines d'articles scientifiques sont parus, documentant l'histoire naturelle et le traitement de la phénylcétonurie ;

> un essai clinique d'un nouveau test diagnostic fut sponsorisé par le Bureau des Enfants des États-Unis et à partir de septembre 1965, le dépistage néonatal de la phénylcétonurie a été établie comme une politique publique de santé par trente-deux états aux États-Unis. Dans vingt-cinq de ces états, le dépistage était obligatoire. Ainsi, dans un très court laps de temps, une maladie génétique rare a été impliquée dans un système divers de prise en charge médicale et politique. » (Lindee, 2000, p. 239)

La phénylcétonurie devient alors l'exemple emblématique de la maladie génétique : une maladie rare, héréditaire, monogénique, mendélienne. Mais dans les années qui vont suivre, plusieurs découvertes vont remettre en cause cette simplicité du modèle de la maladie génétique et ce sont ces découvertes que nous allons analyser maintenant à la lumière des transformations subies par OMIM. L'évolution d'OMIM depuis sa création est effectivement caractérisée par deux transformations majeures : d'une part l'extension du système de classification des phénotypes, d'autre part l'extension du système de classement des objets référencés.

1.3. Les principales transformations d'OMIM – bouleversement de l'hérédité mendélienne et extension du système de classification aux maladies non mendéliennes

1.3.1. Le système de classement des phénotypes : un classement en fonction du mode de transmission de la maladie

A chaque entrée OMIM correspond un code de six chiffres. Le premier chiffre de la série désigne le type d'hérédité auquel se rapporte le phénotype ou le locus concerné, comme le montre le tableau ci-dessous (Tableau 3). En fait, aux débuts du catalogue, le code ne comportait que quatre chiffres : le premier chiffre désignait le type d'hérédité, les chiffres suivants désignaient le classement alphabétique des phénotypes référencés par le nom le plus classiquement utilisé. À partir de la troisième édition du catalogue, en 1975, étant donné l'explosion du nombre de données obtenues grâce aux techniques de la biologie moléculaire, on est passé à un code à 5 chiffres en rajoutant un 0 supplémentaire aux codes à quatre chiffres des entrées

préexistantes ; puis le code a comporté six chiffres par le même système à partir de la neuvième édition (1990).

Le premier chiffre de ce code à six chiffres permet donc de distinguer différentes catégories d'hérédité. Rappelons brièvement que l'hérédité mendélienne est le mode de transmission d'un caractère qui suit les lois de Mendel. On peut résumer les découvertes de Mendel ainsi : un caractère héréditaire peut se présenter sous deux formes différentes (les allèles) et au moment de sa formation, un organisme hérite deux facteurs héréditaires (les gènes) pour chaque caractère. L'effet d'un facteur dominant masque l'effet d'un facteur récessif sur le caractère final. Les deux facteurs se séparent lors de la formation des gamètes et les paires de facteurs se séparent de façon indépendante les unes des autres. On peut donc distinguer quatre formes d'hérédité mendélienne selon que les caractères en question se trouvent sur des chromosomes autosomiques ou sexuels : autosomique dominant, autosomique récessif, lié à l'X et lié à l'Y. L'hérédité mendélienne des chromosomes sexuels est un peu particulière : il y a bien des phénomènes de dominance ou de récessivité pour le chromosome X, mais il n'y en a pas pour le chromosome Y qui, par définition, ne se trouve jamais en paire. On emploie indifféremment les termes d'hérédité mendélienne et d'hérédité monogénique (le caractère phénotypique est causé par un seul gène). En 1966, lorsque MIM avait pour sous-titre, « Catalogue des phénotypes autosomiques dominants, autosomiques récessifs et X liés », la classification ne comportait que trois catégories (catégorie 1 à 3, voir Tableau 3), s'accordant ainsi au titre principal du catalogue qui semble ne traiter que les phénotypes d'hérédité mendélienne. À partir de 1994 (correspondant à la onzième édition du catalogue), trois changements importants surviennent (voir Tableau 3) :

- La base de données s'élargit aux phénotypes liés à l'Y (création de la catégorie 4)
- La base de données s'élargit aux phénotypes mitochondriaux (création de la catégorie 5)
- La catégorie 6 est créée (« phénotypes ou loci autosomiques ») pour se substituer aux catégories 1 (« phénotypes ou loci autosomiques dominants ») et 2 (« phénotypes ou loci autosomiques récessifs »). Toute entrée créée après mai 1994 qui aurait été classée dans les catégories 1 ou 2 est donc classée dans la catégorie 6.

1	Phénotypes ou loci autosomiques dominants (avant 1994)
2	Phénotypes ou loci autosomiques récessifs (avant 1994)
3	Phénotypes ou loci liés à l'X
4	Phénotypes ou loci liés à l'Y (catégorie créée en 1994)
5	Phénotypes ou loci mitochondriaux (catégorie créée en 1994)
6	Phénotypes ou loci autosomiques (catégorie créée en 1994)

Tableau 3 : Évolution du système de classification des phénotypes dans OMIM

Comment comprendre ces changements ? L'inclusion des phénotypes liés à l'Y et aux phénotypes mitochondriaux traduit la prise en compte de modes d'hérédité non mendéliens. La création de la catégorie 6 obéit à une autre logique et répond à un double objectif. Il s'agit d'une part de prendre en compte la relativité des concepts de dominance et de récessivité, deux concepts phares du modèle de la maladie monogénique mendélienne. Par ailleurs, il s'agit de ne plus réserver OMIM aux seules maladies monogéniques mais de prendre en compte les phénotypes complexes polygéniques.

1.3.2. La découverte et l'inclusion des modes d'hérédité non mendélienne

1.3.2.1. L'inclusion des phénotypes liés à l'Y, une continuité logique

En 1966, lors de la première édition du catalogue, McKusick avait inclus des phénotypes liés à l'X qui étaient susceptibles de suivre un mode de transmission mendélien. De ce point de vue, il était logique d'inclure également les phénotypes liés à l'Y. Remarquons cependant, comme on l'a souligné plus haut, que contrairement aux phénotypes autosomiques ou aux phénotypes liés à l'X, il n'y a pas de phénomène de dominance ou de récessivité pour les phénotypes liés à l'Y.

1.3.2.2. L'inclusion des phénotypes mitochondriaux, un problème plus délicat

Les phénotypes mitochondriaux sont des phénomènes particulièrement complexes à comprendre et la raison de leur inclusion dans OMIM n'est pas évidente. Les phénotypes mitochondriaux sont des anomalies de la chaîne respiratoire mitochondriale. Or les protéines de la chaîne respiratoire mitochondriale ont une double origine génétique : elles peuvent provenir soit du génome nucléaire, soit du génome mitochondrial. Tous les modes d'hérédité peuvent donc être observés : une hérédité dite « maternelle » (car lors de la formation du zygote, l'œuf fécondé hérite toujours des mitochondries de la mère) pour le génome mitochondrial et tous les modes d'hérédité mendélienne pour le génome nucléaire. Par ailleurs, un grand nombre de délétions ne sont tout simplement pas héritées mais apparaissent *de novo*. Enfin, la répartition au hasard de l'ADN mitochondrial lors des multiples divisions cellulaires des cellules eucaryotes explique la présence, dans la même cellule, d'ADN mitochondrial normal et d'ADN mitochondrial muté *de novo*. Cette coexistence, appelée hétéroplasmie, explique la grande variabilité dans l'expression clinique de la plupart des syndromes mitochondriaux à la fois d'un patient à l'autre et chez le même patient. En effet, les manifestations pathologiques apparaissent dès lors qu'un certain niveau d'ADN muté s'accumule dans le tissu concerné. Cette répartition aléatoire est aussi à l'origine des atteintes multiviscérales des cytopathies mitochondriales : s'il arrive que les conséquences phénotypiques d'une maladie mitochondriale ne concernent qu'un seul organe, il est beaucoup plus fréquent que plusieurs organes n'appartenant pas au même système anatomoclinique soient touchés ; on parle d' « association illégitime » pour décrire ce phénomène.

Les phénotypes mitochondriaux soulèvent donc deux problèmes. Tout d'abord, ils ne font pas seulement appel aux mécanismes de l'hérédité mendélienne, on pourrait dire qu'ils représentent une forme d'hérédité « non conventionnelle ». Du point de vue de la classification, cette difficulté à identifier l'hérédité des phénotypes mitochondriaux se marque par une dispersion des entrées liées à ces phénotypes à l'intérieur de la classification. En dehors de la catégorie spécifiquement créée pour les mutations d'origine mitochondriale et qui correspond à la catégorie 5, on trouve des renvois à des entrées qui répertorient des mutations qui suivent les lois de l'hérédité

mendélienne classique et qui sont répertoriées soit dans la catégorie 1 ou 2 (lorsque leur découverte est antérieure au 15 mai 1994), soit dans la catégorie 6 (lorsque leur découverte est postérieure à cette date). Par ailleurs, la présence des mitochondries dans toutes les cellules, les phénomènes d'hétéroplasmie et la double origine génétique des cytopathies mitochondriales rendent extrêmement difficile toute corrélation génotype/phénotype. Même le pronostic des maladies mitochondriales est très variable : bien que l'évolution soit généralement rapidement dramatique, certains patients ont vu leurs symptômes régresser ou même complètement disparaître. L'âge auquel apparaît la maladie est également extrêmement variable : bien que dans la plupart des cas les symptômes apparaissent avant la première année, certains ne se déclarent qu'à partir de l'âge adulte.

Du point de vue de la classification OMIM, l'inclusion des phénotypes mitochondriaux ne peut avoir qu'une seule explication : après avoir tenté de ne prendre en compte que les phénotypes pathogènes de type mendélien, McKusick et les membres du *Johns Hopkins Institute* ont été graduellement amenés à inclure dans leur catalogue toute variation génétique qui contribue d'une façon ou d'une autre à l'apparition d'un phénotype pathogène. Cette extension d'OMIM a entraîné non seulement l'inclusion des phénotypes mitochondriaux mais aussi l'inclusion des anomalies chromosomiques, celle des maladies épigénétiques et la numérotation des variants alléliques.

1.3.2.3. Les anomalies chromosomiques : présentes mais invisibles

L'inclusion des aberrations chromosomiques dans la classification OMIM est tout à fait significative de l'extension d'OMIM à des formes de désordres génétiques d'hérédité non mendélienne. En effet, comme nous l'avons décrit précédemment, les anomalies chromosomiques, bien qu'elles puissent être dues pour quelques cas rares à des mutations transmises et héritées sur un mode mendélien, sont souvent dues à des mutations spontanées ou à des accidents de cassure chromosomique lors de la formation du zygote. Les anomalies chromosomiques sont donc des anomalies qui ne font pas spécifiquement appel aux modes de transmission de l'hérédité mendélienne et pourtant McKusick reconnaît les avoir inclus dans OMIM. Autre sujet

d'étonnement : bien que McKusick reconnaisse explicitement avoir ouvert OMIM aux aberrations chromosomiques, il n'y a pas de catégorie spécifique créée pour elles. Par conséquent, les anomalies chromosomiques sont « présentes, mais invisibles » dans OMIM : elles sont dispersées un peu dans chaque catégorie, sans que nous parvenions à comprendre ce qui motive la répartition de telle anomalie dans telle catégorie. Par exemple le syndrome de Down apparaît dans la catégorie 1 : pourquoi mettre le syndrome de Down dans la catégorie des phénotypes autosomiques dominants ?

Si nous ne parvenons pas à comprendre la façon dont sont classées les anomalies chromosomiques dans OMIM, nous disposons tout de même de quelques éléments pour comprendre les raisons qui ont poussé McKusick à les inclure dans la classification. En dehors d'une volonté d'ouvrir la classification à tout type de variation génétique liée à un phénomène pathogène, l'inclusion des aberrations chromosomiques est en lien avec la cartographie génétique : ce sont les techniques cytologiques utilisées pour observer telle aberration chromosomique qui ont permis de faire avancer la cartographie du génome humain et de comprendre la fonction des gènes délétés. En effet, le pourcentage de recombinaison entre deux gènes donne une indication sur leur proximité physique dans le génome, qui peut conduire à l'identification du gène « responsable » de la maladie. C'est ainsi que le gène lié à la maladie de Duchenne a été localisé sur le chromosome X puis identifié en 1986. De façon générale, tout type d'anomalie chromosomique peut mettre les généticiens sur la piste de gènes candidats responsables de la maladie ou de chromosomes candidats. On voit bien ici la nature profonde d'OMIM : c'est une base de donnée créée pour des chercheurs dans un contexte de recherche biomédicale pure et sans rapport aucun avec la prise en charge clinique ou thérapeutique.

1.3.3. La remise en question des concepts de dominance et de récessivité dans l'hérédité mendélienne

Si la création des catégories 4 et 5 traduit l'ouverture d'OMIM à des modes d'hérédité non mendélienne, la création de la catégorie 6 obéit bien à une double logique : d'une part, la reconnaissance du caractère relatif des notions de dominance et de récessivité, concepts propres à l'hérédité mendélienne et d'autre part, l'ouverture d'OMIM aux maladies polyfactorielles, c'est-à-dire à l'hérédité

polygénique. Nous allons d'abord commencer par pointer les différentes difficultés que posent les concepts de dominance et de récessivité : problème du niveau d'analyse du phénotype et concept de semi-dominance, influence des facteurs environnementaux, problème de la pénétrance incomplète, exemple de dominance spécifique de l'hétérozygote.

1.3.3.1. Problème du niveau d'analyse du phénotype et de la semi-dominance

C'est ainsi que McKusick justifie l'abandon de la distinction entre phénotypes autosomiques dominants et phénotypes autosomiques récessifs :

> « Par ailleurs, la distinction est seulement relative – le fait d'être dominant ou récessif dépend parfois seulement du niveau auquel on analyse le phénotype. Par exemple dans plusieurs enzymopathies des globules rouges, le déficit est autosomique récessif mais se traduira probablement par une variation électophorétique chez l'hétérozygote, c'est-à-dire que le trait est dominant ou au moins intermédiaire. » (McKusick, 2007, p. 588)

Il s'agit ici de distinguer les niveaux d'analyse dans l'étude d'un phénotype – en l'occurrence, celui de l'anémie falciforme (autrement appelée « drépanocytose » ou *sickle-cell disease/anemia* en anglais). L'anémie falciforme est une maladie des globules rouges : ceux-ci ont une forme particulière « en faucille », qui entraîne différentes conséquences pathogènes : anémie, ictère, hépatosplénomégalie, susceptibilité aux infections, crises douloureuses, crises vaso-occlusives, accidents thromboemboliques, etc. Elle a une prévalence très importante dans les populations noires : le fait d'avoir des globules rouges en forme de faucille accroît en effet la résistance au paludisme et c'est ce qui explique la persistance de cette mutation qui a longtemps constitué un facteur de sélection favorable. Ainsi, à l'heure actuelle, aux États-Unis, un enfant sur cinq cent naît atteint d'une anémie falciforme parmi la population afro-américaine.

Il existe plusieurs formes d'anémie falciforme : chacune implique à un moment donné une substitution Glu-Val (l'acide glutamique est remplacé par la valine) sur l'une des chaînes bêta qui forment l'hémoglobine. Généralement les anémies falciformes sont considérées comme des maladies autosomiques récessives, dans la

mesure où il faut posséder les deux gènes mutés pour voir se développer les symptômes de la maladie. Soit HbA l'allèle normal et HbS l'allèle muté : les individus HbA/HbA sont associés à un phénotype non pathogène, les individus hétérozygotes HbA/ HbS ont un phénotype non pathogène et seuls les individus HbS/ HbS ont un phénotype pathogène, c'est-à-dire seuls les individus HbS/ HbS sont anémiés. Si cependant on regarde ces individus au niveau de la protéine, on observe cette fois-ci trois phénotypes possibles : les homozygotes pour l'allèle normal ne produisent qu'une seule forme d'hémoglobine, la forme normale, les homozygotes pour l'allèle muté produisent une seule forme d'hémoglobine, la forme pathogène, et les hétérozygotes produisent les deux types d'hémoglobine en des proportions variables selon les individus. Du point de vue de la production de l'hémoglobine, l'anémie falciforme est donc dominante ou représente du moins ce qu'on appelle un phénomène de « semi-dominance » : il suffit d'être hétérozygote pour voir apparaître une protéine mutée et anormale. C'est un exemple de problème de seuil pathologique : en fonction des mutations, de l'influence de gènes modificateurs ou de facteurs environnementaux, l'expressivité de la protéine mutée varie et il faut atteindre un certain seuil de cette expressivité pour que la maladie apparaisse. Or, il est pour l'instant impossible de déterminer le seuil d'expressivité qu'atteindra la protéine mutée sur la seule base de l'identification des mutations génétiques.

De la même façon, on dit que la plupart des maladies autosomiques dominantes ne sont que semi-dominantes. En effet, la présence d'un allèle muté suffit certes à provoquer l'apparition de la maladie mais on ne trouve pas les mêmes symptômes pathogènes chez l'hétérozygote cliniquement malade et chez l'homozygote : les effets des deux allèles mutés s'additionnent et donnent une maladie beaucoup plus sévère à l'état homozygote qu'à l'état hétérozygote.

1.3.3.2. Problème de l'intervention des facteurs environnementaux

Les notions de dominance et de récessivité sont problématiques parce qu'elles reposent sur le postulat d'une relation de déterminisme entre la présence d'un allèle et l'apparition d'un phénotype, alors même que d'autres facteurs entrent en ligne de compte comme les facteurs environnementaux. Ainsi, dans des conditions

« normales » d'oxygénation, l'individu hétérozygote ne sera pas anémié, alors qu'un individu hétérozygote pour l'anémie falciforme qui se trouve en altitude et s'expose à des conditions d'oxygénation raréfiée a de grandes chances de développer les symptômes aigues de la maladie.

C'est le même type de difficulté qui se pose avec le favisme ou déficit en glucose 6 phosphate déshydrogénase (qui est une enzyme des globules rouges). Cette affection est génétique et liée au sexe, car elle provient d'un gène anormal porté par le chromosome sexuel X. Transmis par les mères, ce déficit atteint essentiellement les garçons. C'est une particularité génétique d'un enfant normal. En effet, l'enfant porteur de ce déficit est un enfant normal qui ne sera malade de son déficit, que s'il subit des agressions extérieures dites « oxydatives », causées par l'ingestion de certains aliments ou médicaments. À cet égard, le déficit en G6PD pourrait ne même pas être considéré comme une maladie puisque l'individu ne développe aucun symptôme, tant qu'il n'entre pas en contact avec l'agent oxydatif. Il est donc nécessaire d'observer un régime d'exclusion de certains aliments (comme les fèves, d'où le nom de « favisme ») et de prendre certaines précautions vis-à-vis de quelques médicaments susceptibles de causer une anémie parfois très brutale. Mais dans un environnement où le régime est contrôlé, l'enfant ne sera pas cliniquement malade, même s'il peut être considéré sur le plan moléculaire comme porteur d'un déficit. C'est un cas d'école en santé publique car la prévention joue un rôle majeur, et passe par l'information adéquate des patients porteurs de ce déficit, de leur entourage familial et scolaire, ainsi que des médecins.

1.3.3.3. Problème de la pénétrance incomplète et de l'expressivité variable

La pénétrance est la proportion d'individus possédant un génotype donné qui exprime le phénotype correspondant. Lorsque la pénétrance est incomplète, cela signifie que des individus porteurs d'un même génotype n'exprimeront pas le même phénotype. Le problème de la pénétrance incomplète ne doit pas être confondue avec celui de l'expressivité variable. L'expressivité variable désigne le fait que pour une même maladie, il n'y a pas de corrélation parfaite entre la présence de certaines mutations et l'apparition de certains symptômes. Autrement dit, le même allèle

pathogène peut causer des symptômes cliniques variables d'un individu à un autre. Ces deux phénomènes peuvent survenir au sein d'une même famille (variation intrafamiliale) et dans la même fratrie. Dès lors, la notion de dominance et de récessivité est en un sens mise à mal : la transmission des mutations peut bien être héréditairement dominante, si le phénotype clinique a une pénétrance incomplète ou une expressivité variable au sein d'une même famille, il semble difficile de parler de dominance à proprement parler.

Prenons l'exemple de la rétinite pigmentaire. Le terme de « rétinite pigmentaire » recouvre un ensemble assez varié d'affections bilatérales rétiniennes qui conduisent à la perte des photorécepteurs et donc la plupart du temps à la cécité. Il y a différents modes de transmission de la rétinite pigmentaire : autosomique dominant, autosomique récessif, liée à l'X ou mitochondrial. Dans un tiers des cas, la maladie est sporadique. La forme dominante (représentant environ 35% des cas) est associée à des mutations dans une vingtaine de gènes impliqués dans le métabolisme rétinien (une mutation sur chacun de ces gènes pouvant entrainer une forme dominante autosomique de la rétinite pigmentaire), en conséquence de quoi les formes cliniques de la maladie sont très hétérogènes.

Mais une étude de 2011 (Maubaret *et al.,* 2011) montre que même lorsqu'on s'intéresse à un seul gène (en l'occurrence le gène PRFP8), les manifestations cliniques de la maladie sont hétérogènes d'un individu à un autre (expressivité variable) et les études familiales montrent des cas de pénétrance incomplète (les individus sont porteurs de la mutation mais ne développent pas la maladie). Dans cette étude, il s'agissait de comparer les corrélations génotype / phénotype parmi les dix individus de deux familles (chaque famille étant affectée par une mutation différente du gène PRPF8). Dans la première famille, tous les individus porteurs de la mutation ont exprimé à un moment donné de leur vie des signes d'héméralopie (perte de la vision nocturne, généralement considérée comme un des premiers signes de la maladie) mais à des âges très différents les uns des autres, vers l'âge de dix ans pour les plus jeunes et à la fin de la trentaine pour les plus vieux (phénomène d'expressivité variable). L'analyse de la seconde famille est encore plus intéressante. On y retrouve la même variabilité clinique intra-familiale, mais aussi un cas de pénétrance incomplète où le patient était porteur de la même mutation que le reste de la famille mais n'avait

aucun symptôme clinique ou paraclinique (aucun symptôme infraclinique n'avait été révélé par ailleurs par les différents examens ophtalmologiques).

1.3.3.4. Problème de la dominance autosomique spécifique de l'hétérozygote

Rappelons encore une fois le concept de dominance et de récessivité. Une maladie est transmise sur un mode dominant, lorsque l'allèle muté détermine le phénotype à l'état d'hétérozygotie. Une maladie est transmise sur le mode récessif, lorsque la présence d'un seul allèle muté n'entraîne aucune variation phénotypique chez l'hétérozygote, la maladie n'apparaissant que chez les porteurs de deux allèles mutés. Une équipe de chercheurs canadiens (Raymond, 1998) a cependant découvert un nouveau mécanisme dans une forme héréditaire du glaucome à angle ouvert résultant d'une mutation dans le gène codant la myocyline (OMIM 601652). Dans cette maladie, lorsque le porteur est homozygote pour deux copies mutées du gène TIGR (*trabecular meshwork-inducible glucocorticoid response protein*), il est cliniquement sain et ne développe pas la maladie. En revanche, lorsque il est porteur hétérozygote, la maladie se développe. Un tel phénomène est appelé « dominance autosomique de l'hétérozygote » et peut s'expliquer sur le plan moléculaire. La protéine en question agit en effet sous la forme de multimère : quand il y a un mélange de protéines normales et de protéines mutées, le complexe perd sa fonctionnalité et le patient développe la maladie. Mais lorsque toutes les protéines sont mutées, le complexe semble parvenir à se former (bien que cette formation diffère de la formation du complexe à partir de protéines normales), au point que le complexe redevient fonctionnel. Cet exemple illustre à la fois le caractère relatif des notions de dominance et de récessivité et les difficultés de la prédiction du phénotype à partir du génotype, dès lors que l'on ignore les phénomènes moléculaires intermédiaires qui entrent en compte dans la détermination d'un phénotype.

1.3.4. L'ouverture d'OMIM aux maladies polygéniques

La naissance de la catégorie 6 ne répond pas seulement à l'effacement d'une distinction absolue entre maladies mendéliennes dominantes et maladies mendéliennes récessives. Elle traduit aussi l'ouverture d'OMIM aux maladies

polygéniques et l'inclusion des variants alléliques dans le système de classification. Nous allons aborder ce point à travers la description des transformations du système des symboles.

1.4. Les principales transformations d'OMIM – d'un classement des maladies à un classement des variants alléliques

L'ouverture d'OMIM aux maladies polygéniques s'est traduite par l'inclusion des variants alléliques, qu'ils soient fortement corrélés à l'apparition d'une maladie mendélienne ou bien qu'ils constituent un polymorphisme fréquent dans une maladie multifactorielle. Cette inclusion des variants allélique est une vraie révolution dans la classification des maladies génétiques : il ne s'agit plus seulement de répertorier des désordres mendéliens à pénétrance forte touchant majoritairement un seul gène ; il s'agit de répertorier toute variation qui pourrait être significative. Ce changement de perspective que nous avons appréhendé via le système de numérotation dans la classification OMIM se révèle aussi quand on étudie l'histoire du système de symboles qui complète la référence de chaque entrée.

Le point majeur de cette transformation du système de symboles a lieu en 1990. À partir de cette date, on a systématiquement séparé les entrées qui désignent des phénotypes et les entrées qui désignent les gènes qui causent ces mêmes phénotypes. L'extension de la classification (on ne classe plus seulement des maladies mais aussi des gènes de façon distincte) a été définitivement consacrée avec le changement de sous-titre en 1994 (« Catalogue des gènes humains et des désordres génétiques »). Comme l'explique McKusick lui-même (McKusick, 2007, p. 589), cette séparation entre les phénotypes et les gènes est nécessaire pour rendre compte de plusieurs phénomènes qui compliquent la corrélation phénotype/génotype et qui ont contribué à remettre en cause le modèle de la maladie monogénique : « un phénotype, plusieurs gènes » (hétérogénéité génétique), « un gène, plusieurs phénotypes » (pléiotropie ou hétérogénéité phénotypique) et enfin « plusieurs allèles, une même maladie » (hétérogénéité allélique).

1.4.1. Le système des symboles : un classement en fonction de la nature de l'objet référencé

Qu'est-ce que le système de symboles ? Devant le code d'une entrée, on trouve un certain nombre de symboles qui déterminent en grande partie la nature de l'objet référencé dans cette entrée :

- Un astérisque (*) indique un gène dont la séquence est connue
- Un signe dièse (#) indique une entrée descriptive, généralement celle d'un phénotype, et ne représente pas un locus unique. La raison pour laquelle on utilise le signe dièse est donnée dans le premier paragraphe de l'entrée. Les discussions qui cherchent à déterminer si un ou plusieurs gènes sont reliés au phénotype sont dans d'autre(s) entrée(s) indiquées dans le premier paragraphe.
- Un signe (+) indique que l'entrée contient la description d'un gène de séquence connue et d'un phénotype.
- Un signe pourcentage (%) indique que l'entrée décrit un phénotype mendélien confirmé ou un locus responsable d'un phénotype mendélien pour lequel les mécanismes moléculaires sous-jacents sont inconnus.
- L'absence de symbole peut indiquer soit que la base mendélienne du phénotype décrit n'a pas été clairement établie, bien qu'elle soit suspectée, soit qu'il est difficile de distinguer ce phénotype d'une autre entrée.
- Un accent circonflexe (^) signifie que l'entrée n'existe plus car elle a été retirée de la base de données ou déplacée à une autre entrée.

Un exemple de ce type de classement des entrées est donné dans la partie « statistiques » de la base de données et nous avons reproduit ce classement pour le 7 janvier 2013 dans le tableau ci-dessous (Tableau 4).

OMIM Statistics for January 7, 2013

Number of Entries

	Autosomal	X-Linked	Y-Linked	Mitochondrial	**Total**
* Gene with known sequence	13370	651	48	35	14104
+ Gene with known sequence and phenotype	124	4	0	2	130
# Phenotype description, molecular basis known	3371	271	4	28	3674
% Mendelian phenotype or locus, molecular basis unknown	1627	133	5	0	1765
Other, mainly phenotypes with suspected mendelian basis	1765	125	2	0	1892
Total	20257	1184	59	65	**21565**

Tableau 4 : Classement par symbole des différents objets référencés sur OMIM. Capture d'écran du 7 février 2013 (Source site web d'OMIM, partie statistiques).

Ce système de symboles a également connu de profondes transformations historiques. En 1966, pour la première édition, seuls les phénotypes et non les gènes ou leurs loci étaient classés dans le catalogue. Deux types d'entrées étaient alors distinguées : les entrées avec astérisque (*) et les entrées sans. Les entrées avec astérisque désignaient d'une part les phénotypes dont le mode de transmission héréditaire était considéré comme clairement identifié et certain, et d'autre part, les phénotypes que l'on parvenait clairement à délimiter des autres. Par opposition, les phénotypes sans astérisque étaient soit des phénotypes dont l'origine mendélienne était moins claire, ou bien des phénotypes qu'on arrivait mal à distinguer de phénotypes assez proches. À partir de 1990, on a systématiquement séparé les entrées qui désignent des phénotypes et les entrées qui désignent les gènes qui causent ces mêmes phénotypes. L'extension de la classification a été définitivement consacrée avec le changement de sous-titre en 1994 (« Catalogue des gènes humains et des désordres génétiques »).

1.4.2. Le concept de variant allélique

Les phénotypes ne sont donc pas les seuls objets que classe OMIM, les variants alléliques sont également numérotés : ils portent alors un code à dix chiffres, les six premiers chiffres ne changent pas et désignent le gène, les quatre derniers sont spécifiques de l'allèle en question. Ainsi pour le principal gène impliqué dans la phénylcétonurie (PAH : Phenylalanine Hydroxylase, MIM*612349, voir Figure 4), qui code l'enzyme phénylalanine hydroxylase, responsable de la transformation de la phénylalanine en tyrosine, on trouve à l'heure actuelle 67 variants alléliques mentionnés et portant les numéros *612349.0001 à *612349.0067 dans la rubrique variants alléliques. À ces variants alléliques est associée une formule entre crochets qui désigne le type de mutation spécifique du variant allélique concerné. Par exemple, pour le variant allélique *612349.0002 PAH, on trouve la formule associée : [PAH, ARG408TRP] qui signifie que sur le gène PAH, une délétion du codon 408 a entraîné la délétion de l'acide aminé phénylalanine.

612349

PHENYLALANINE HYDROXYLASE; PAH

Allelic Variants (Selected Examples):

Number	Phenotype	Mutation	dbSNP
.0001	PHENYLKETONURIA	PAH, IVS12DS, G-A, +1	-
.0002	PHENYLKETONURIA	PAH, ARG408TRP	[rs5030858]
.0003	PHENYLKETONURIA	PAH, LEU311PRO	[rs62642936]
.0004	PHENYLKETONURIA	PAH, GLU280LYS	[rs62508698]
.0005	PHENYLKETONURIA	PAH, ARG111TER	[rs76296470]
.0006	PHENYLKETONURIA	PAH, ARG261GLN	[rs5030849]
.0007	PHENYLKETONURIA	PAH, ARG252TRP	[rs5030847]
.0008	PHENYLKETONURIA	PAH, EX3DEL	-
.0009	PHENYLKETONURIA HYPERPHENYLALANINEMIA, NON-PKU, INCLUDED	PAH, MET1VAL	[rs62514891]
.0010	PHENYLKETONURIA	PAH, ARG158GLN	[rs5030843]
.0011	PHENYLKETONURIA	PAH, ARG243TER	[rs5030846]
.0012	PHENYLKETONURIA	PAH, PRO281LEU	[rs5030851]
.0013	PHENYLKETONURIA	PAH, TYR204CYS	[rs62514927]
.0014	PHENYLKETONURIA	PAH, ARG243GLN	[rs62508588]

Figure 4 : Liste (partielle) des variants alléliques pour le gène de la phénylcétonurie (*612349). Source : http://omim.org/entry/612349. (Capture d'écran du 7 février 2013).

Il faut préciser que si 67 variants alléliques sont indiqués ici, il existe en fait 567 variants alléliques répertoriés aujourd'hui dans OMIM. En effet, il est précisé sur le site d'OMIM que la liste des variants donnés pour un gène donné est indicative (et non exhaustive). A quelles conditions inclut-on alors un variant allélique ? Dans son article

de 2007 que nous avons précédemment cité, McKusick identifie neuf critères possibles. Ainsi, on choisira d'inclure en priorité :

- les premières mutations reliées à la maladie et qui ont été identifiées pour un gène donné
- toute mutation qui a une fréquence particulièrement élevée
- toute mutation reliée à un phénotype distinct qui n'a pas été représenté auparavant dans le texte descriptif de la mutation ou du gène
- des mutations d'intérêt historique, comme la mutation spécifique associée à la famille ou à la population où le phénotype a été la première fois décrite : par exemple, la mutation dans le gène CLCN1 (MIM*118425.0006) dans la famille du Dr Thomsen qui a le premier décrit la maladie de Thomsen (MIM#160800)
- toute mutation associée à une distribution géographique ou ethnique spécifique
- toute mutation qui apparaît par un mécanisme mutagène bien spécifique comme la conversion génétique
- toute mutation qui entraîne un phénotype pathogène par un mécanisme pathogénétique spécifique
- les mutations qui sont associées à des modes de transmission qui peuvent être soit autosomique dominant soit autosomique récessif et qui sont des mutations différentes du même gène : par exemple les formes récessives (MIM*139250.0005) et dominantes (MIM*139250.0007) de la déficience isolée de l'hormone de croissance qui sont dues à des mutations alléliques différentes dans le gène de l'hormone de croissance (GH1)
- les polymorphismes dont on a démontré qu'ils sont fortement corrélés avec une maladie particulière (ces variants alléliques sont aussi appelés « gènes de susceptibilité ») : par exemple le polymorphisme Y402H du facteur complémentaire H (CFH [MIM*134370.0008]) est fortement corrélé à une certaine forme de dégénérescence maculaire (MIM#603075).

Le fait d'inclure tout type de variant allélique, qu'il soit fortement corrélé à l'apparition d'une maladie mendélienne ou qu'il constitue un polymorphisme fréquent dans une maladie multifactorielle, indique véritablement une révolution dans la classification des maladies génétiques : il ne s'agit plus seulement de répertorier des désordres mendéliens à pénétrance forte touchant majoritairement un seul gène ; il

s'agit de répertorier toute variation du matériel génétique qui pourrait être significative dans l'apparition d'une maladie.

1.4.3. Hétérogénéité génétique, hétérogénéité allélique et pléiotropie

Comme le dit lui-même McKusick, il s'agissait de rendre compte de plusieurs phénomènes qui compliquaient la corrélation phénotype/génotype : *« La pratique des entrées séparées pour les gènes et les phénotypes a été initiée en grande partie pour résoudre les problèmes un phénotype-plusieurs gènes et un gène-plusieurs phénotypes. »* (McKusick, 2007, p. 589). Aux deux phénomènes cités par McKusick, nous en ajoutons un troisième, le phénomène « plusieurs allèles, une même maladie » (hétérogénéité allélique). Nous allons à présent les décrire un par un.

1.4.3.1. Le phénomène « un gène-plusieurs phénotypes » (pléiotropie ou hétérogénéité phénotypique)

Ce phénomène s'explique par les découvertes de la biologie moléculaire. Un gène est un morceau d'ADN, c'est-à-dire une combinaison des quatre nucléotides qui forment les bases du code génétique. Ces quatre nucléotides sont : A pour adénine, C pour cytosine, G pour guanine et T pour tyrosine. Pour être exprimé, c'est-à-dire pour conduire à la production d'une protéine, un gène doit passer par deux phases : une phase de transcription et une phase de traduction. La phase de transcription permet le passage de l'ADN nucléaire à un ARN messager qui sort du noyau et permet l'initiation de la phase de traduction qui a lieu dans le cytoplasme. Cette phase de traduction est aussi une phase de lecture : chaque triplet de nucléotides (codon) code en effet pour un acide aminé ; c'est la traduction de l'ensemble des codons qui permet l'assemblage d'acides aminés et la production d'une protéine.

Si un morceau d'ADN nucléaire donnait lieu à une copie identique d'ARN messager et à la production d'une seule protéine, les mécanismes de la génétique seraient assez simples à comprendre. Mais en réalité, l'ADN nucléaire d'un gène est composé de parties codantes (ce qu'on appelle des exons) et de parties non codantes (les introns) : après qu'il y a eu production d'un pré-ARN messager, alors qu'il est encore dans le noyau, ce pré-ARN messager est épissé, c'est-à-dire débarrassé des

introns. Mais il existe plusieurs types d'épissage possibles, on parle donc d' « épissage alternatif » et ces différents épissages vont donner lieu à différentes protéines et donc à différentes conséquences phénotypiques à long terme. C'est un effet de pléiotropie qui est encore compliqué par d'autres processus post-traductionnels (qui ont lieu après la traduction). Le résultat de ce phénomène est qu'un seul gène peut être à l'origine de plusieurs phénotypes ou plus exactement, qu'un seul gène peut causer des maladies différentes et des variations phénotypiques importantes pour une même maladie. Un bon exemple de ce genre de difficultés concerne le gène lamine A/C : ce gène est susceptible de donner plus de onze maladies différentes de la progéria à une forme de la maladie Charcot-Marie Tooth. Ce problème est d'ailleurs encore compliqué par la notion de gènes mosaïques auquel nous avons fait brièvement allusion.

1.4.3.2. Le phénomène « un phénotype, plusieurs gènes » (hétérogénéité génétique)

Ce phénomène se traduit par le fait qu'un même phénotype peut être causé par des mutations affectant différents gènes. L'explication en est assez simple : pour assurer la bonne marche d'une chaîne métabolique ou d'un processus physiologique complexe, plusieurs protéines et plusieurs gènes entrent en ligne de compte, un phénomène connu aussi sous le nom de redondance fonctionnelle. La « faillite » de composants différents d'une même chaîne métabolique peut ainsi entraîner le même phénotype apparent. Un bon exemple de ce problème est le syndrome du QT long : les onze formes répertoriées de ce syndrome correspondent à des mutations dans différents gènes. On pourrait aussi citer le cas des rétinopathies pigmentaires dans lesquelles plus de quarante gènes seraient parfois impliqués.

1.4.3.3. Le phénomène « plusieurs allèles, une même maladie » (hétérogénéité allélique)

Ce phénomène rend compte du fait qu'une maladie peut être due à plusieurs allèles morbides du même gène. En fonction de l'allèle morbide, la conséquence phénotypique sera plus ou moins importante et la maladie plus ou moins grave. Un

bon exemple est la phénylcétonurie et le gène PAH : plus de cinq cent mutations ont été répertoriées en lien avec l'apparition de la maladie.

Ces différents éléments permettent de comprendre l'évolution nécessaire du système de référence de la classification McKusick :

> « Depuis le début, nous avions toujours à l'esprit le gène derrière le phénotype. Pendant la majorité des quinze années du catalogue cependant, il n'était que rarement possible de savoir si un phénotype donné était en fait causé par des mutations dans un seul gène ou pourrait être causé par des mutations dans deux ou plusieurs gènes et il était en général impossible de dire si deux ou plusieurs phénotypes différents étaient dus à des mutations dans un même gène. Depuis 1990, la règle est devenue de créer des entrées séparées pour les phénotypes et pour les gènes qui présentent des mutations causant ces phénotypes. » (McKusick, 2007, p. 589)

Autrement dit, les objectifs de la classification de McKusick étant avant tout de corréler génotype et phénotype, à partir du moment où l'équation un gène/un phénotype ne peut plus être respectée, il est nécessaire de changer le mode de classement et de chercher à corréler gène et phénotype par un système de renvois et de références croisées. Dès lors, la classification des gènes et des désordres génétiques reflète véritablement l'état de la recherche biomédicale à un temps *t* ; ce qui explique la prise en compte de gènes qui n'ont pas encore été clonés mais dont on suppose qu'ils sont corrélés à un désordre génétique et la prise en compte de phénotypes dont on n'a pas déterminé l'origine génétique, mais dont on pense par l'analyse d'arbres familiaux qu'ils ont une grande chance d'être héréditaires ou d'avoir une composante génétique.

1.4.4. Deux conséquences de l'ouverture d'OMIM aux maladies polyfactorielles

Cette ouverture de la classification aux maladies polyfactorielles et aux variants génétiques posent deux questions : évolue-t-on vers une reclassification moléculaire de la maladie ? Combien de maladies génétiques y a-t-il aujourd'hui ?

1.4.4.1. Vers une classification moléculaire des maladies génétiques ?

Dès lors que les outils de la génétique furent suffisamment précis pour distinguer plusieurs phénotypes à partir de leurs caractéristiques moléculaires, les tentatives de reclassement des maladies génétiques en fonction de ces marqueurs sont devenues inévitables. Le passage d'une classification phénotypique à une classification moléculaire est lourd de conséquences. Au dix-neuvième siècle, autour de la structure de l'hôpital s'est mise en place une pratique de la médecine occidentale fondée sur la distinction clinique et phénotypique des maladies. C'est en regroupant un ensemble de symptômes anatomo-physiologiques qu'on distinguait des entités pathogènes différentes. Le passage d'une distinction clinique à une distinction moléculaire modifie considérablement les familles de maladies avec lesquelles les médecins avaient l'habitude de s'orienter dans la nosologie.

> « La classification actuelle des maladies et le diagnostic médical sont les conséquences directes d'une généralisation inductive qui applique le rasoir d'Occam. Cette approche scientifique a longtemps bien servi les cliniciens dans leurs efforts pour établir des modèles de syndromes qui rationalisent le nombre de phénotypes à considérer. Par ailleurs, étant donné la faiblesse des informations quantitatives disponibles pour raffiner et décomposer plus en profondeur ces phénotypes, l'exercice du diagnostic était intrinsèquement limité tout en étant maniable pour le praticien. Ces limitations diagnostiques deviendront cependant très bientôt une note de bas de page d'un intérêt historique. À présent que la fin du séquençage du génome humain est devenue une réalité et que nous disposons d'un ensemble croissant de données transcriptomiques, protéomiques et métaboliques dans le domaine de la santé et de la maladie, nous sommes dans une situation unique dans l'histoire de la médecine, celle de pouvoir définir la maladie humaine de façon précise, unique et non équivoque, avec une sensibilité et une spécificité optimales. En théorie, cette caractérisation moléculaire précise des maladies humaines devraient nous permettre de comprendre les bases de la susceptibilité à la maladie et des influences de l'environnement, d'offrir une explication aux manifestations phénotypiques différentes d'une même maladie, de définir le pronostic de la maladie avec une plus grande précision et

d'individualiser le traitement médical pour une efficacité thérapeutique optimale. » (Loscalzo *et al.*, 2007, p. 124)

La plupart des commentateurs ne sont pas loin de partager la vision de Loscalzo : l'avènement d'une reclassification moléculaire des maladies est vécue comme un progrès exceptionnel en termes de prise en charge de l'individualité de chaque patient, aussi bien sur le plan diagnostique que sur le plan thérapeutique. Nous nous permettons cependant une courte parenthèse pour signaler que cet enthousiasme est modéré par certains sociologues de la médecine, qui estiment qu'une telle reclassification risque d'exclure comme non malades des patients qui présentent des symptômes d'une maladie sans en avoir la caractérisation moléculaire ou d'inclure comme malades des patients qui n'ont pas de symptômes mais qui ont la mutation (Miller *et al.,* 2005) . Mais au-delà d'une discussion sur les bienfaits ou les inconvénients d'une reclassification moléculaire des maladies, un tel mouvement signe aussi la fin d'une longue controverse : deux camps se sont traditionnellement affrontés sur la meilleure manière de comprendre et de distinguer les phénotypes pathogènes ; les *lumpers* (en français les « rassembleurs ») et les *splitters*[8] (en français les « diviseurs ») : les *lumpers* considèrent que les différents phénotypes cliniques sont des manifestations différentes d'une seule entité clinique sous-jacente et que les différences apparentes ne sont pas aussi importantes que les similarités. Les *splitters* pensent que ces différentes manifestations représentent des variations importantes, voire des entités cliniques sous-jacentes différentes qui doivent être reconnues comme telles et traitées différemment. Pour donner un exemple simple et peut-être un peu caricatural des influences que peut avoir une telle prise de position, on peut dire que dans le domaine de la recherche du cancer, les *lumpers* auront tendance à chercher le remède du cancer tandis que les *splitters* auront tendance à chercher des remèdes différents pour une centaine de cancers différents. Finalement, il semblerait que, si on considère l'ensemble des classifications des maladies génétiques, les *lumpers* comme les *splitters* aient en quelque sorte gagné puisque il existe à la fois des phénotypes distincts qui ont pour cause la même mutation et des phénotypes

[8] La controverse entre *lumpers* et des *splitters* ne concerne pas que la classification des maladies génétiques, elle peut s'appliquer à toutes les taxinomies ou à toute tentative de classification (McKusick, 1969).

semblables qui ont pour cause des mutations différentes dans plusieurs gènes. Néanmoins, la controverse resurgit fréquemment de façon ponctuelle sur des familles de maladies précises[9].

1.4.4.2. Combien de maladies génétiques ? Combien de gènes de la maladie ?

Étant donné les incertitudes qui continuent cependant à planer sur la reclassification moléculaire des maladies génétiques (à partir de combien de mutations différentes pour le même gène considère-t-on que l'on parle de deux maladies différentes ? Faut-il mettre sur le même plan hétérogénéité génétique et hétérogénéité allélique ?), il est extrêmement difficile de dénombrer les maladies génétiques. C'est un problème qui est fréquemment relevé par les commentateurs. Ainsi, dans le manuel de génétique médicale du Collège des Enseignants et des Praticiens de Génétique Médicale, sous la direction du professeur Arnold Munnich, aucun chiffre précis n'est donné mais on trouve la remarque suivante :

> « Le nombre des maladies génétiques connues est souvent surévalué par confusion entre le nombre de gènes impliqués et le nombre de phénotypes ; certains gènes sont à l'origine de plusieurs maladies, certaines maladies mendéliennes sont liées à plusieurs gènes. » (Collège national des enseignants et praticiens de génétique médicale, 2004)

Les phénomènes de pléiotropie, d'hétérogénéité génétique et d'hétérogénéité allélique rendent donc difficile, voire impossible une estimation du nombre de maladies génétiques existantes. Dans la rubrique « Aide » d'OMIM, ce problème est clairement abordé : pour obtenir le nombre de maladies génétiques (associant donc un phénotype avec un gène), il faut prendre en compte toutes les catégories, sauf celles assorties du préfixe « * », c'est-à-dire sauf les gènes dont la séquence est connue : on obtient alors un nombre de 7 461 maladies pour le 7 février 2013 ; mais on est face aux mêmes ambiguïtés sur la précision de ce nombre.

[9] Pour une approche plus détaillée et consensuelle de ce problème, voir (Berenson, 2004). Pour quelques exemples de phénotypes précis qui causent des difficultés de classifications, voir (Traboulsi *et al.*, 2006) pour la maladie de Leber et voir (Shea, 2005) pour l'autisme.

1.5. Le concept de maladie génétique aujourd'hui – entre extension et dissolution

1.5.1. Les maladies génétiques dans OMIM – bilan et perspectives

Qu'est-ce que l'analyse d'OMIM nous apprend ? Si le but de la classification n'a pas changé depuis son origine (il s'agit toujours de corréler génotype et phénotype, et de parvenir à une meilleure délimitation des maladies génétiques), la présentation et le contenu d'OMIM ont beaucoup évolué pendant les quarante dernières années. Un peu grossièrement, on peut résumer cette évolution comme le passage d'une classification qui ne concernait que les maladies monogéniques mendéliennes simples à une classification qui concerne toute variation génétique, quel que soit son mode de transmission et quelle que soit la précision avec laquelle elle est corrélée à un phénotype, à partir du moment où elle est susceptible d'intervenir à quelque stade que ce soit dans l'apparition d'un phénotype pathogène. Cette évolution peut donner à OMIM des airs de « bricolage ». Cependant, elle semble rétrospectivement inéluctable pour rendre compte de la complexité des phénomènes biologiques en jeu et de la rapidité des avancées de la recherche biomédicale. Même les phénotypes mendéliens, considérés traditionnellement comme des traits simples, semblent désormais être considérés comme des maladies complexes (Scriver et Waters, 1999 ; Dipple et McCabe, 2000a, 2000b) et la frontière entre maladies monogéniques simples et maladies multifactorielles complexes semble de plus en plus ténue (Badano et Katsanis, 2002 ; Rousseau et Laflamme, 2003). C'est ce qu'exprime Victor McKusick, qui va jusqu'à évoquer la nécessité d'ouvrir OMIM aux « maladies épigénétiques » :

> « Étant donné le rôle des facteurs génétiques dans la pathogénèse, j'ai comme d'autres trouvé utile dans les premières étapes du développement de la génétique médicale de diviser les maladies – de façon plutôt arbitraire il est vrai, en maladies mendéliennes, chromosomiques et multifactorielles. La nature arbitraire de cette classification n'enlève rien à son utilité ; toutes les classifications sont d'une certaine manière artificielle. Pour être cohérent avec la perspective garrodienne, toute variation génétique, y compris celles qui contribuent seulement au caractère multifactoriel des traits complexes, et

> peut-être aussi à celui des traits chromosomiques (variations génomiques) doit être cataloguée dans OMIM. Ces vingt dernières années, il est devenu évident que la variation épigénétique est devenue une quatrième classe étiopathogénétique majeure, en particulier dans les cancers. » (McKusick, 2007, p. 596)

McKusick évoque ici l'ouverture à une autre classe de maladie, les maladies épigénétiques et cela, bien qu'aucune date précise ne soit donnée pour l'ouverture à cette classe de maladie et bien qu'aucun principe d'organisation (création d'une catégorie particulière) ne soit suggéré pour s'en accommoder. Qu'est-ce qu'une maladie épigénétique ? On parle de variations épigénétiques pour décrire des variations qui influencent l'expression des gènes et qui ne sont pas codées comme une information génétique, bien qu'elles soient transmises par l'un des deux parents. Plusieurs mécanismes entrent en jeu dans les variations épigénétiques. Un des mécanismes les plus connus est l'empreinte génomique parentale. Les empreintes génomiques parentales sont des phénomènes physiologiques qui conduisent dans des conditions normales à l'expression d'une seule des deux copies d'un gène, soit la copie paternelle, soit la copie maternelle. On parle de « gènes soumis à une empreinte maternelle », lorsque ces gènes sont inactifs sur le chromosome maternel et de « gènes soumis à une empreinte paternelle » lorsque les gènes sont inactifs sur le chromosome paternel. L'empreinte génomique parentale résulte d'une marque épigénétique apposée sur certaines régions du génome des cellules germinales, qui inactive l'expression des gènes situés dans ces régions. La forme de cette marque est en général une méthylation de l'ADN ou des histones qui va permettre des modulations de l'activité de transcription du ou des gènes (c'est le cas le plus fréquent, on parle de *clusters* de gènes) sous la dépendance de cette marque. L'exemple (non pathogène) d'empreinte génomique parentale le plus célèbre est celui de l'inactivation du chromosome X. Pour éviter que les filles produisent deux fois plus de protéines codées par le chromosome X, un des deux chromosomes est inactivé très tôt au cours de l'embryogénèse. Les gènes soumis à empreinte parentale jouent en général un rôle dans la croissance fœtale et dans le développement postnatal : une anomalie de l'empreinte génomique parentale conduit donc le plus souvent à des phénotypes de malformations, à des anomalies de la croissance et à un risque plus élevé de

développer certaines tumeurs chez l'enfant ou chez l'adulte. Ce qui complique la compréhension des phénotypes liés à des anomalies des gènes soumis à l'empreinte parentale, c'est que le mécanisme de l'empreinte n'est pas actif dans les mêmes cellules et au même moment du développement d'un organisme. Il y a d'ailleurs différents types de mécanismes épigénétiques. On distingue entre autres :

(1) Un mécanisme épigénétique par régulation de la région du promoteur : le promoteur est une séquence d'ADN qui contrôle le début de la transcription d'un ou plusieurs gènes. Pour que le promoteur soit activé et que la transcription puisse commencer, il faut que des facteurs de transcription puissent accéder à la région du promoteur. Un mécanisme d'empreinte génomique parentale qui méthyle l'ADN de la région du promoteur va entraîner une conformation de la chromatine qui empêche l'accès des facteurs de transcription à la région du promoteur et donc qui bloque la transcription des gènes sous la dépendance de ce promoteur.

(2) Un mécanisme épigénétique par la régulation des *enhancers* et des *silencers* : on appelle «*enhancers* » des séquences d'ADN qui stimulent la transcription d'un ou plusieurs gènes et « *silencers* » des séquences d'ADN qui inhibent la transcription d'un ou plusieurs gènes. En contrôlant la régulation des *enhancers* et des *silencers* par des facteurs répresseurs ou par des phénomènes de méthylation de l'ADN, l'empreinte génomique parentale contrôle l'expression d'un grand nombre de gènes.

Les anomalies des mécanismes de l'empreinte génomique parentale peuvent donc être à l'origine de syndromes phénotypiques clairement identifiés comme le syndrome de Prader-Willi ou de phénomènes tumoraux[10]. Nous ne voulons pas davantage détailler ce point : il nous semble en revanche que l'ouverture aux variations épigénétiques est tout à fait dans la lignée des transformations d'OMIM que nous avons présentées jusqu'ici et qu'elle témoigne des bouleversements qu'a connus le concept de maladie génétique ces quarante dernières années.

Nous pouvons à présent décrire la situation paradoxale dans laquelle se trouve la génétique des maladies aujourd'hui. Le concept de maladie génétique semble s'être

[10] Il suffit en effet que l'expression d'un gène suppresseur de tumeur soit inhibée par un mécanisme d'empreinte génomique parentale pour que le patient ait un risque très élevé de développer des tumeurs ; c'est un de ces mécanismes sur le chromosome paternel 7 qui conduit ainsi à la leucémie aigüe myéloblastique.

étendu bien au-delà de son sens originel, au point de potentiellement concerner toute maladie. Dans le même temps, cette extension conceptuelle s'accompagne d'une remise en question paradoxale du concept de maladie génétique : si toute maladie est génétique, que veut donc dire ce concept ?

1.5.2. La généticisation des maladies

Le concept de maladie génétique s'est étendu bien au-delà de son sens originel. En effet, si aux débuts d'OMIM, le concept de maladie s'appliquait à des maladies rares, monogéniques, héréditaires, mendéliennes dont l'exemple paradigmatique est la phénylcétonurie, ce terme désigne également aujourd'hui des maladies communes, non héréditaires, non mendéliennes et polygéniques – au point que toute maladie peut être considérée comme génétique, un phénomène que la sociologue Abby Lippman a appelé la première une « généticisation » (Lippman, 1991) :

> « La généticisation désigne le processus actuel par lequel priorité est donnée à la recherche des variations de séquences d'ADN qui différencient les individus les uns des autres et par lequel un fondement héréditaire est attribué à la plupart des maladies, des comportements et des variations physiologiques (ce qui inclut aussi bien la schizophrénie et la pression artérielle que la capacité des enfants à rester assis pendant qu'ils regardent la télévision ou la capacité des adultes à arrêter de fumer). » (Lippman, 1994, p. 13)

Lorsqu'Abby Lippman a forgé ce terme, c'était dans le contexte bien précis d'une réflexion critique sur le rôle des technologies génétiques (dépistage prénatal, test génétique, etc.) dans notre perception des maladies. La définition de la généticisation est donc dès son origine non seulement une définition porteuse de connotations intrinsèquement négatives mais aussi une définition complexe, mêlant plusieurs critères : (1) c'est un processus sociologique qui peut affecter et être affecté par les perceptions culturelles et sociales locales (les mêmes technologies dans un environnement culturel différent auraient pu avoir des conséquences différentes et inversement, la généticisation des maladies a des conséquences sur la perception culturelle et sociale de la maladie dans un environnement donné), (2) c'est un processus lié à l'expansion des techniques de biologie moléculaire et au

développement d'applications médicales de ces techniques, (3) c'est un processus qui revient à différencier les individus à partir de leurs différences génétiques et à attribuer de façon croissante une explication génétique aux maladies et aux comportements. Notons d'ailleurs une ambiguïté dans la définition proposée par Lippman qui semble mélanger « fondement héréditaire » et différence moléculaire, alors même qu'une anomalie moléculaire n'est pas forcément héritée.

Mais cette définition de la généticisation va encore gagner en complexité, en raison du succès considérable de ce terme, qui a été repris aussi bien par les philosophes (voir, par exemple, Hoedemaekers et ten Have, 1997) que par les sociologues (voir, par exemple, Nelkin et Lindee, 1996), au point d'avoir encore perdu en clarté. Par exemple, dans l'ouvrage de Nelkin et Lindee, où les deux auteures analysent les discours sociaux, médiatiques et populaires sur le concept de gène, la généticisation est assimilée à « l'essentialisme génétique ». L'essentialisme génétique désigne le processus par lequel l'ADN n'est plus seulement considéré comme une entité biologique, mais comme un symbole, une métaphore de l'essence humaine. En conséquence de quoi, l'essentialisme génétique « réduit le soi à une entité moléculaire, restreignant les êtres humains dans toute leur complexité sociale, historique et morale à leurs gènes » (Nelkin et Lindee, 1995, p. 2). Là encore, la connotation négative domine, la généticisation se rapproche d'un discours idéologique dépassant très largement l'explication des maladies.

À l'opposé de cette généticisation considérée d'un point de vue global comme l'expansion symbolique d'un discours essentialiste sur le gène, plusieurs sociologues ont proposé des analyses locales et neutres d'exemples précis du processus de la généticisation des maladies. Cette fois-ci la généticisation est comprise comme un processus de molécularisation (détachée du concept d'hérédité) et comme un changement dans l'explication de ces maladies. C'est ainsi qu'Ann Kerr a analysé la reconstruction de la mucoviscidose comme une maladie génétique et la reclassification de l'agénésie congénitale des canaux déférents comme une forme minime de la mucoviscidose du fait de l'implication des mutations du gène CFTR dans ces deux maladies (Kerr, 2000, 2004, 2005). Adam Hedgecoe a décrit la façon dont la classification du diabète a évolué pour reposer en partie sur des différences génétiques entre diabète de type 1 et diabète de type 2, sous l'impulsion de deux

articles écrits en 1976 et en 1978 par le diabétologue Andrew Cudworth mais aussi sous l'impulsion de différentes institutions. L'objectif avoué d'Adam Hedgecoe est de soustraire le concept de généticisation à une définition trop large, trop vague et fondamentalement négative pour l'intégrer à une analyse sociologique plus dépouillée :

> « La généticisation est un processus social potentiellement porteur d'impacts éthiques, légaux et politiques, mais dans cet article, ce qui m'intéresse davantage c'est d'attirer l'attention sur la façon dont la recherche sur la généticisation doit être portée par des sociologues. J'espère que cet article remet en cause l'idée que la généticisation ne puisse pas être examinée ou évaluée par les sciences sociales. La généticisation peut se produire de façon subtile et inattendue, sans le recours à des technologies de haute volée ou au séquençage de génomes entiers. Ce qui est nécessaire, c'est une recherche empirique précise sur la façon dont la généticisation prend place et est portée dans des contextes aussi variés que possible ». (Hedgecoe, 2002, p. 23)

Pour résumer, la généticisation n'admet pas de définition simple et peut désigner à peu près n'importe quoi entre un processus sociologique, issu de discours médiatiques sur le pouvoir des gènes et une « molécularisation » de l'explication médicale, avoir des connotations négatives ou en être dénuée, s'appliquer à toutes les maladies ou désigner un processus local de changement d'explication pour certaines maladies (Hedgecoe, 1998). C'est pourquoi nous décidons d'adopter ici la définition, aussi neutre que possible, qu'Alan Hedgecoe propose dans son article concernant la généticisation du diabète :

> « Une solution est d'adopter une définition « toute nue » de la généticisation, en affirmant simplement qu'en médecine il y a généticisation lorsqu'une maladie est reliée à un brin d'ADN spécifique. Une telle définition permet de se focaliser sur des périodes, des individus ou des technologies particuliers tout en soulignant ces aspects de la science qui ont rendu une telle « relation » possible. Il y a de nombreuses objections à cette définition : le concept de « relation » serait vague, tout comme celui de « brin » d'ADN et on viderait le concept original de Lippman de toute sa charge critique au risque qu'il ne perde son pouvoir persuasif. Cette définition est délibérément ambiguë : l'idée

> de « relation » couvre tout de la corrélation entre certains allèles et certains phénotypes au séquençage et au clonage des gènes. De la même façon, en exigeant que cette relation soit un brin d'ADN identifié, mon concept de généticisation est fermement ancré dans la génétique moléculaire : cette définition est plus spécifique que la suggestion qu'une maladie a un composant héréditaire, mais elle est plus large que la nécessité pour un gène spécifique d'être séquencé. » (Hedgecoe, 2002, p. 8‑9)

Nous partageons le point de vue de Hedgecoe pour trois raisons. Premièrement, il nous semble qu'avant de décider si la généticisation est un phénomène souhaitable ou pas, ou de chercher à comprendre les origines du phénomène, en somme avant de proposer une interprétation de la généticisation, il faut d'abord adopter un sens neutre et restrictif qui permette de caractériser le phénomène. Par ailleurs, sa définition en terme de « relation » entre une maladie et un brin d'ADN spécifique nous paraît particulièrement adaptée à ce que nous avons décrit de l'évolution d'OMIM. Le caractère héréditaire mendélien d'une maladie n'est plus nécessaire à sa caractérisation comme maladie génétique, le fait qu'elle soit associée à des variants alléliques suffit. Enfin la définition d'Hedgecoe permet de caractériser deux phénomènes : *la généticisation d'une maladie particulière*, c'est-à-dire la façon dont une maladie spécifique va être liée à un brin d'ADN donné[11] et *la généticisation des maladies* en général comme la tendance à considérer toute maladie comme une maladie génétique. C'est principalement ces deux phénomènes que nous chercherons à interpréter dans la suite de ce travail.

1.5.3. La dissolution du concept de maladie génétique

La généticisation des maladies interroge le concept de maladie génétique : si toute maladie est génétique, que veut donc dire le concept de maladie génétique ? Mais cette question est d'autant plus prégnante que les problèmes du concept de maladie génétique ne sont pas seulement des problèmes d'extension. Ce n'est pas seulement le fait d'inclure des maladies non héréditaires, non mendéliennes ou non

[11] Pour la description de la généticisation de l'anémie falciforme, on pourra se reporter par exemple à (Hedgecoe, 2003 ; Anne Kerr, 2005 ; Kerr, 2000). Pour d'autres exemples de généticisation, voir pour le diabète et la schizophrénie : (Hedgecoe, 2002 ; Turney and Turner, 2000).

monogéniques qui met en tension l'unité du concept de maladie génétique, c'est aussi le fait que les concepts clés qui sous-tendaient le modèle de la maladie génétique (la notion de « maladie monogénique » ou de « maladie mendélienne », la distinction entre dominance et récessivité) ont été remis en question, déstabilisant les fondements de notre compréhension de la maladie génétique.

Ainsi, si l'on reprend l'exemple de la phénylcétonurie, force est de constater, en suivant Charles Scriver, d'une part que l'explication contemporaine qui en est donnée est bien plus complexe que celle proposée en 1963 et d'autre part, que la phénylcétonurie et les maladies monogéniques mendéliennes en général ne sont pas des maladies simples. (Scriver et Waters, 1999 ; Scriver, 1995, 2007). L'explication de 1963 était la suivante : un gène muté et alors inconnu, dont on avait démontré la transmission mendélienne récessive et dont on supposait qu'il codait pour l'enzyme de la phénylalanine hydroxylase responsable de la conversion de la phénylalanine en tyrosine, entraînait un défaut en phénylalanine hydroxylase, une augmentation de la phénylalanine dans le sang et un retard intellectuel important. En 2007, ce schéma un gène - une enzyme - une maladie est singulièrement remis en question par une importante hétérogénéité allélique (il y a plus de cinq cents variants alléliques du gène PAH associés au développement de la phénylcétonurie), une importante hétérogénéité génétique (des mutations dans le gène codant l'enzyme BH4, co-facteur de l'enzyme phénylalanine hydroxylase peuvent entraîner une augmentation du taux de phénylalanine, avec des symptômes similaires à la phénylcétonurie), une pénétrance incomplète (il est très difficile de prédire le phénotype en fonction du génotype, certains individus avec les mêmes variants alléliques mutés auront des phénotypes différents, certains exhibant même des phénotypes sains). La phénylcétonurie est-elle une maladie simple ? Si on ne s'intéresse qu'aux défauts présentés par le gène PAH, la maladie est mendélienne. Si l'on s'intéresse à l'apparition d'un excès de phénylalanine dans le sang, la maladie est multifactorielle au sens où il faut prendre en compte à la fois le régime en phénylalanine et les mutations. Le phénotype final est d'ailleurs un phénotype fondamentalement complexe puisqu'il dépend a) du phénotype de l'enzyme (le taux de phénylalanine circulant dans le plasma) qui lui-même est lié à la vitesse de dégradation de la protéine, b) mais aussi du phénotype métabolique (la façon dont la phénylalanine s'accumule dans le plasma) qui

lui-même est affecté par le transport de la phénylalanine, (c) et enfin du phénotype cognitif (retard mental) qui dépend du transport de la phénylalanine à travers la barrière hémato-encéphalique. C'est cette complexité que la figure ci-dessous représente (Figure 5) :

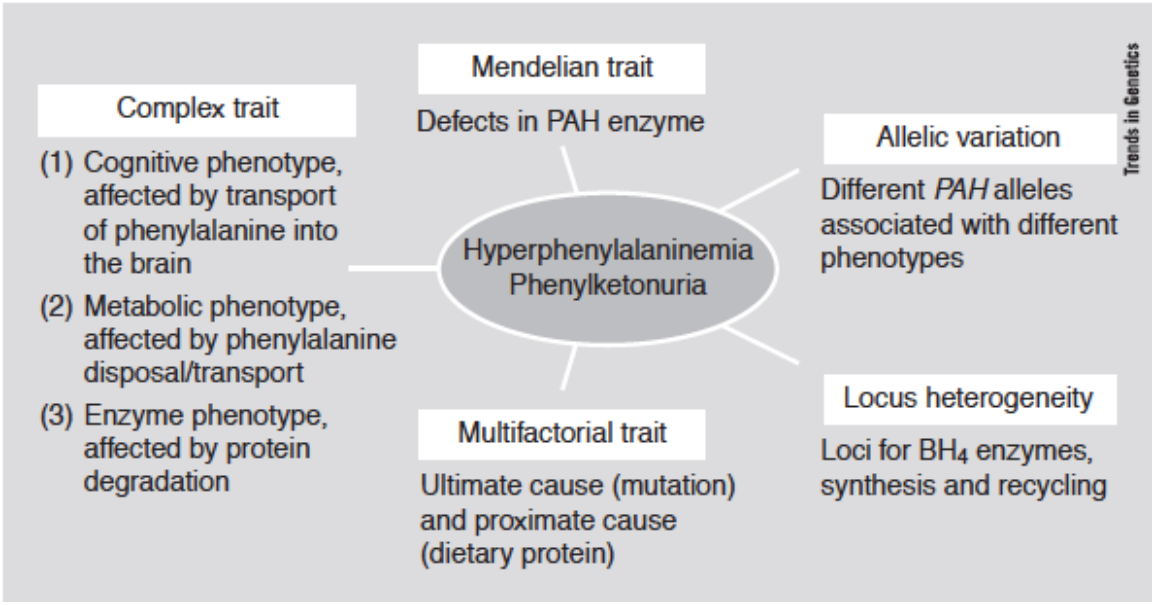

Mendelian trait
Defects in PAH enzyme
Complex trait
(1) Cognitive phenotype, affected by transport of phenylalanine into the brain
(2) Metabolic phenotype, affected by phenylalanine disposal/transport
(3) Enzyme phenotype, affected by protein degradation
Hyperphenylalaninemia
Phenylketonuria
Allelic variation
Different PAH alleles associated with different phenotypes
Locus heterogeneity
Loci for BH4 enzymes, synthesis and recycling
Multifactorial trait
Ultimate cause (mutation) and proximate cause (dietary protein)
Trends in Genetics

Figure 5 : Facteurs influençant le phénotype de la phénylcétonurie (Scriver et Waters, 1999, p. 269)

Pour ajouter à cette complexité, le gène codant l'enzyme BH4 se comporte par ailleurs comme un gène modificateur, en dehors des cas où il est muté aboutissant à un excès de phénylalanine plasmatique sans qu'il y ait de mutations dans le gène PAH. En effet, il semble qu'en fonction de leur génotype sur le locus codant l'enzyme BH4, les patients atteints de la phénylcétonurie (ceux qui ont un gène PAH muté n'aboutissant pas à l'absence de phénylalanine hydroxylase mais à la production d'une phénylalanine hydroxylase majoritairement dysfonctionnelle et qui présentent des phénotypes dits modérément sévères), voient leur symptômes améliorés par l'administration de l'enzyme BH4. Celle-ci permettrait en effet une stabilisation de l'enzyme phénylalanine hydroxylase dysfonctionnelle (Blau *et al.,* 2011). La somme de ces complexités permet de comprendre en quoi le modèle de la maladie monogénique et mendélienne, où la corrélation entre génotype et phénotype est supposée être simple et où la maladie était supposée être presque entièrement déterminée par le gène, est en fait un mirage. Ces différents phénomènes (relativité de la distinction

entre dominance et récessivité, hétérogénéité allélique, hétérogénéité génétique, pénétrance incomplète, hétérogénéité phénotypique, gènes modificateurs) ont été mis en évidence pour un nombre considérable de maladies monogéniques, inspirant à la plupart des généticiens la conclusion suivante :

> « La distinction entre maladies « monogéniques » (les maladies mendéliennes « simples ») et maladies polygéniques (les maladies complexes) est devenue de moins en moins claire au fur et à mesure que l'on a réalisé son caractère artificiel et qu'elle correspondait davantage à une représentation de la perception humaine qu'à la réalité biologique. Nous voudrions affirmer que pour une maladie soi-disant monogénique, il y a un gène qui peut être majoritairement responsable de la pathogénèse avec un ou plusieurs gènes modificateurs hérités de façon indépendante et qui influencent le phénotype. De l'autre côté, pour une maladie complexe, aucun gène individuel ne semble avoir de primauté et la maladie résulte de l'interaction de deux ou de plusieurs paires d'allèles indépendamment héritées, très probablement influencées par ailleurs par des gènes modificateurs additionnels. **La conséquence de ce cadre conceptuel c'est que la « maladie monogénique » n'existe pas**[12]. En d'autres termes, il n'y a pas de distinction absolument claire entre maladies mendéliennes simples et maladies complexes : les maladies génétiques représentent un continuum qui va d'une influence minimale d'un gène majeur par quelques gènes modificateurs à une influence croissante partagée par de multiples gènes. » (Dipple et McCabe, 2000a, p. 47)

Deux points majeurs nous intéressent dans cette situation. Le premier est le lien entre la fragilisation du concept de la maladie monogénique mendélienne et la relativité de la distinction entre maladies monogéniques et maladies polygéniques. C'est parce que les maladies monogéniques ne sont pas simples que cette distinction entre maladies monogéniques et maladies polygéniques, autrefois claire, est devenue brouillée et cela d'autant plus que de nouveaux mécanismes polygéniques ont émergé. Dans les années soixante, les maladies polygéniques étaient comprises comme des maladies communes, complexes où plusieurs gènes, chacun ayant un effet mineur, interagissent avec des facteurs environnementaux pour déterminer le phénotype final.

[12] C'est nous qui soulignons.

Mais se sont forgés depuis d'autres concepts, tels que celui de maladie « oligogénique » : cette notion un peu floue, recouvre l'idée qu'il y a « quelques » gènes majeurs, influencés par des gènes modificateurs qui déterminent le phénotype de la maladie :

> « Récemment, il a été prouvé que de nombreuses maladies qui étaient initialement caractérisées comme monogéniques sont soit causées, soit modulées, par l'action d'un petit nombre de loci. Ces maladies sont décrites comme des maladies « oligogéniques », un concept qui évolue de façon continue et qui prend en compte un large spectre de phénotypes qui ne sont ni monogéniques ni complexes. Contrairement aux traits polygéniques – dont on pense qu'il résulte d'interactions mal comprises entre de nombreux gènes et l'environnement, ces maladies oligéniques restent avant tout génétique dans leur étiologie, mais nécessitent l'action synergique d'allèles mutants à un petit nombre de loci. De nombreux exemples indiquent qu'il existe un continuum entre les maladies mendéliennes classiques et les traits complexes. La position d'une maladie donnée le long de ce continuum dépend de trois variables principales : la contribution marquée d'un locus majeur au phénotype, le nombre de loci impliqués et l'importance de la participation de l'environnement. » (Badano et Katsanis, 2002, p. 780)

Autrement dit, la maladie oligogénique est un pas supplémentaire dans l'effacement de la distinction absolue entre maladie monogénique et maladie polygénique : elle conduit à l'élaboration d'un « continuum génétique » de la maladie, qui est l'autre notion que nous allons à présent développer.

1.5.4. La notion de continuum génétique, révélateur des ambiguïtés de la génétique médicale

La notion de continuum génétique des maladies est particulièrement intéressante : fille de la généticisation mais aussi de la fragilisation du concept de maladie génétique, elle incarne les ambiguïtés et les paradoxes de la génétique des maladies aujourd'hui. On la retrouve chez de nombreux auteurs de la littérature biomédicale contemporaine (Badano et Katsanis, 2002 ; Collège national des

enseignants et praticiens de génétique médicale, 2004 ; Dipple et McCabe, 2000 ; Khoury, 2003 ; Rousseau et Laflamme, 2003). Nous avons choisi d'analyser l'ambiguïté de ce concept en nous intéressant à la façon dont le Collège des Enseignants et des Praticiens de génétique médicale représentent ce continuum (Figure 6). Précisons que le manuel publié par le CNEPGM en 2004 est supposé représenter un consensus français en génétique médicale et est destiné à des étudiants de l'enseignement supérieur en génétique médicale. Ce n'est donc en aucun cas un ouvrage de vulgarisation où l'on pourrait admettre des simplifications à visée pédagogique.

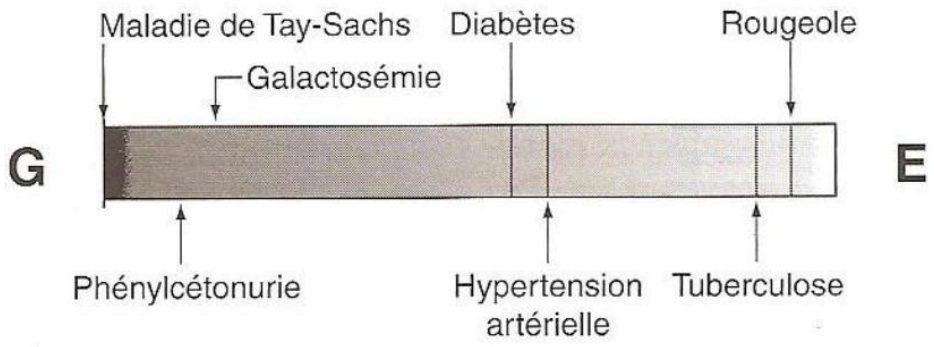


Figure 6 : Diagramme de l'action conjointe des facteurs génétiques (G) et environnementaux (E) sur les malades (Collège national des enseignants et praticiens de génétique médicale, 2004, p. 21)

Cette représentation est accompagnée du texte suivant :

> « Depuis environ quarante ans, les généticiens essaient d'expliquer cette répartition familiale en recherchant une composante génétique dans l'étiologie de ces maladies qui seraient dues à la conjonction de facteurs génétiques et de facteurs d'environnement, d'où leur nom de maladies à déterminisme multifactoriel ou de maladies multifactorielles. L'action conjointe de ces deux grands groupes de facteurs, génétiques et d'environnement, se retrouve en fait dans toutes les maladies. Ce qui diffère d'une maladie à l'autre, c'est la part respective de chacun d'eux : toute maladie peut trouver sa place sur un diagramme tel qu'à gauche se situe le pôle génétique G et à droite, le pôle environnement E. Une maladie donnée sera d'autant plus près du pôle G que la composante héréditaire y sera plus importante. Tout près du pôle G se placent les maladies dont la transmission suit les lois de Mendel. Le milieu peut cependant y jouer un rôle ; en effet,

> chaque fois qu'une maladie de ce type peut être traitée, c'est toujours actuellement par une intervention portant sur le milieu. À l'extrémité droite du diagramme se trouvent les maladies infectieuses ; le rôle de la composante héréditaire y est classiquement faible, mais ceci est actuellement remis en question pour certaines de ces maladies telles que la tuberculose, la lèpre et la bilharziose. En situation intermédiaire, on trouve les maladies communes déjà citées. » (Collège national des enseignants et praticiens de génétique médicale, 2004, p. 21)

L'idée de continuum est clairement présente aussi bien dans le texte que dans le diagramme et semble au premier coup d'œil concerner toutes les maladies (accréditant ainsi la thèse de la généticisation des maladies). En même temps, le diagramme distingue bien entre trois classes de maladie : la portion noire du diagramme semble désigner des maladies qui seraient entièrement déterminées par des facteurs génétiques, la portion blanche des maladies qui seraient entièrement déterminées par des facteurs environnementaux, tandis que la portion grise (et qui est la seule à présenter un dégradé) semble représenter des maladies « plus ou moins génétiques, plus ou moins environnementales ». Le sens de « génétique » n'est d'ailleurs pas tout à fait clair : les auteurs continuent à parler de variation héréditaire, comme si les maladies à mutation somatique (comme la majorité des cancers par exemple) n'étaient pas prises en compte. Il n'est d'ailleurs pas évident que la portion entièrement noire représente des maladies monogéniques, puisque la phénylcétonurie n'en fait pas partie. Quant aux maladies purement environnementales (représentées en blanc à l'extrême droite du graphique), aucun exemple n'est donné – ce qui ne serait pas étonnant si on admet qu'il existe des maladies non génétiques, mais qui paraît surprenant dans la mesure où la citation précédente souligne l'action conjointe de facteurs environnementaux et génétiques dans toutes les maladies.

La notion de continuum génétique reflète donc la généticisation des maladies et la relativité de la distinction entre maladies monogéniques et maladies polygéniques. Mais elle semble aussi aboutir à l'idée que certaines maladies sont plus génétiques que d'autres et plusieurs auteurs semblent avoir du mal à abandonner cette distinction entre maladies monogéniques et polygéniques.

1.6. Conclusion du chapitre : le paradoxe de la génétique médicale contemporaine

L'objectif de ce chapitre était de comprendre ce que recouvrait le concept de maladie génétique. Parce que l'histoire du concept de maladie génétique est particulièrement complexe, nous avons préféré aborder ce problème en étudiant la façon dont les maladies génétiques sont classées. La Classification Internationale des Maladies ne fournit pas de principes de classement des maladies génétiques, celles-ci étant dispersées en fonction de leur localisation anatomo-clinique ou de leur appartenance à des classes étiologiques bien identifiables. Nous nous sommes donc tournée vers *Online Mendelian Inheritance in Man*, une base de données qui a été lancée sous forme de catalogue papier des désordres mendéliens par Victor McKusick en 1966, à une époque où le modèle de la maladie génétique est incarné par la phénylcétonurie et qui est devenue aujourd'hui la base de données de référence des maladies génétiques, répertoriant tous les phénotypes associés à des variations alléliques.

L'étude des transformations d'OMIM depuis son lancement nous a permis de mettre en évidence deux phénomènes majeurs : d'une part, la remise en question de la simplicité des maladies monogéniques mendéliennes (hétérogénéité allélique, hétérogénéité génétique, hétérogénéité phénotypique, pénétrance incomplète, gènes modificateurs, modes d'hérédités non mendéliens tels que les maladies mitochondriales et les anomalies chromosomiques, relativité des notions de dominance et de récessivité) et d'autre part, l'ouverture aux maladies polygéniques mettant en jeu des mutations somatiques non héritées, des variations alléliques multiples et des maladies épigénétiques.

Cette double transformation aboutit ainsi à une situation paradoxale : alors que le concept de maladie génétique s'est étendu au point de pouvoir désigner toute maladie (un phénomène qui peut être qualifié de généticisation des maladies), l'unité du concept de maladie génétique s'est fragilisée. Cette situation paradoxale s'incarne dans le concept de continuum génétique des maladies qui semble à la fois reconnaître la généticisation des maladies et la relativité de la distinction entre maladies

monogéniques et maladies polygéniques, tout en admettant qu'il existe des maladies « plus génétiques » que d'autres.

À présent que nous avons analysé les problèmes scientifiques que soulève le concept de maladie génétique aujourd'hui, nous allons nous tourner vers la façon dont les philosophes interprètent à la fois la généticisation des maladies et le concept de maladie génétique.

Chapitre 2 : Problèmes philosophiques soulevés par le concept de maladie génétique : généticisation, génocentrisme et problème de la sélection causale

2.1. Définir la maladie génétique, une instance du problème de la sélection causale

2.1.1. Le problème de la sélection causale

À présent que nous avons décrit les problèmes scientifiques posés par le concept de maladie génétique, il s'agit de décrire les stratégies philosophiques les plus communes pour résoudre ce paradoxe de la génétique médicale contemporaine, que nous avons défini comme la remise en cause du concept de maladie génétique et la généticisation des maladies. Il y a bien ici deux questions à résoudre : comment les philosophes interprètent-ils l'affirmation selon laquelle « toute maladie est génétique » et comment définissent-ils le concept de maladie génétique ?

Il est frappant de constater que la plupart des philosophes qui se sont intéressés à cette question (Gifford, 1989, 1990 ; Hesslow, 1984 ; Hull, 1979a, 1979b, 1997 ; Magnus, 2004 ; Smith, 1992, 2001, 2007 ; Wulff, 1984) partagent un même présupposé : tous considèrent que définir la maladie génétique est une instance du problème de la sélection causale. Qu'est-ce que le problème de la sélection causale ? Dans un article de 1983, Germund Hesslow identifie une première formulation de ce problème, parfois aussi appelé « le problème de l'importance relative des causes » dans les écrits de John Stuart Mill lorsque celui-ci cherche à définir la différence entre causes et conditions (Mill, 1869) ou dans le célèbre article de Mackie en 1965, lorsque celui-ci évoque l'exemple de l'incendie d'une maison :

> « Supposons qu'une certaine maison a pris feu, mais que le feu a été éteint avant que le maison ne soit complètement détruite. Les experts enquêtent sur la cause du feu et ils concluent que le feu a été causé par un court-circuit électrique à un certain endroit. Quelle est la force exacte de leur affirmation selon laquelle c'est le court-circuit qui a causé le feu ? Evidemment, les experts ne disent pas que le court-circuit était une condition nécessaire pour que la

> maison prenne feu à ce moment-là ; ils savent parfaitement qu'un court-circuit à un autre endroit de la maison, ou un poêle à mazout allumé qui se serait renversé ou bien d'autres événements encore, s'ils étaient arrivés, auraient pu déclencher le feu. De la même manière, ils ne disent pas que le court-circuit était une condition suffisante pour que la maison prenne feu ; car si le court-circuit avait eu lieu mais qu'il n'y avait pas eu de matériau inflammable à proximité, le feu n'aurait pas pris et même en admettant la présence du court-circuit et du matériau inflammable, le feu n'aurait pas pris, si, par exemple, il y avait eu un système d'extinction automatique efficace juste au bon endroit. Loin d'être une condition nécessaire et suffisante pour le déclenchement du feu, le court-circuit n'était pas – et les experts le savent bien, ni une condition nécessaire, ni une condition suffisante pour que le feu se déclenche. Dans quel sens, alors, dit-on qu'il a causé le feu ? » (Mackie, 1965, p. 245)

Dans cet exemple, il s'agit de fournir une explication causale de l'incendie. En l'occurrence, les experts chargés d'enquêter sur la cause de l'incendie ont choisi de sélectionner une cause comme la cause la plus importante parmi un faisceau de causes distinctes. Ce qui est étonnant, c'est qu'alors que nous acceptons dans la vie quotidienne cette explication causale, nous sommes dans le même temps conscients (« et les experts le savent bien », dit Mackie) que le feu, comme la plupart des événements, a plusieurs causes et que le court-circuit n'est ni une condition suffisante ni une condition nécessaire de l'incendie.

Le problème de la sélection causale est donc distinct du problème de l'inférence causale, aussi appelée « connexion causale », qui consiste à déterminer dans quelles circonstances et par quelle méthode on peut affirmer qu'un événement est l'une des causes d'un autre événement. Comme le dit Germund Hesslow :

> « La plupart des discussions philosophiques sur le problème de la sélection sont fondées sur le présupposé qu'un énoncé causal repose sur deux jugements distincts, chacun correspondant à un problème conceptuel : l'un concerne l'existence d'une relation causale entre deux événements, le « problème de la connexion » et l'autre concerne l'importance relative de ces causes, le « problème de la sélection ». Le problème de la connexion revient à comprendre par quel processus nous déterminons, par exemple, que la

> présence de l'oxygène, la présence de matériel combustible et la présence d'une source de chaleur sont des conditions nécessaires à ce qu'une maison prenne feu. Une fois qu'on a reconnu la pertinence causale de ces conditions, il reste à savoir cependant laquelle de ces conditions était la plus importante dans un cas individuel concret. Nous ne disons pas que le feu était causé par l'oxygène en dépit du fait que nous savons qu'il y a une connexion entre l'oxygène et le feu. » (Hesslow, 1988, p. 13)

Le problème de la sélection causale revient donc à déterminer pourquoi la plupart des explications causales ne prennent pas en compte l'ensemble des causes d'un événement mais en sélectionnent une comme la cause « la plus décisive » :

> « Ainsi, nous sommes confrontés à cette situation qu'un événement normalement a plusieurs, peut-être même un nombre infini de causes, mais que seules certaines d'entre elles sont sélectionnées et citées dans les explications causales. Parfois, nous parlons même de la cause d'un événement qui a pourtant dans les faits plusieurs causes. Pourquoi cela ? Comment est déterminée la sélection de la cause la plus importante parmi l'ensemble complet des conditions causales ? C'est précisément le problème de la sélection causale qui nous intéresse dans ce chapitre. » (Hesslow, 1988, p. 12)

Comment ce problème de la sélection causale qui a été discutée dans d'autres domaines que celui de la médecine (pour un exemple en physique, voir Sober, 1988), s'applique-t-il à la question des maladies génétiques ? Deux présupposés sont partagés par tous les commentateurs du concept de maladie génétique. Le premier présupposé est que les maladies sont des événements mettant en jeu un grand nombre de causes et que les facteurs génétiques et non génétiques interviennent probablement d'une manière ou d'une autre dans toute maladie. Par exemple, même dans le cas de l'empoisonnement à l'arsenic, maladie « environnementale » s'il en est, il y a probablement entre les individus des différences génétiques quant à la façon dont le système immunitaire peut réagir à l'arsenic. Toute maladie est donc le produit d'un faisceau à la fois de causes génétiques et environnementales :

> « En effet, il semble que *tous* les traits sont le résultat d'une interaction entre les facteurs causaux génétiques et environnementaux. Car l'on peut toujours

> citer ou concevoir un changement génétique qui serait pertinent pour la présence ou l'absence d'un trait. De la même façon, l'on peut toujours citer ou concevoir un facteur environnemental avec la même propriété. » (Gifford, 1990, p. 329)

Le deuxième présupposé découle du premier. Lorsque nous définissons une maladie comme génétique, cela ne revient évidemment pas à dire que le gène est le seul facteur causal de la maladie, puisque par principe, il est toujours possible de concevoir une cause environnementale. Cela signifie plutôt que nous résolvons le problème de la sélection causale en faveur des facteurs génétiques : en définissant une maladie comme génétique, nous affirmons que le gène est le facteur causal le plus important – ce qui n'empêche pas de reconnaître l'intervention de facteurs environnementaux. Nous voulons donner ici plusieurs citations pour exemplifier le caractère commun de cette stratégie argumentative parmi les commentateurs :

> « Il est évident que dans la plupart des cas de maladies génétiques, l'application de ce terme reflète le choix de souligner l'importance parmi les facteurs de la composante génétique et de minimiser l'importance des facteurs environnementaux ». (Hull, 1979b, p. 78)

> « Aucun trait n'a pour *seule* cause un facteur génétique. Cela ne peut donc pas être la définition de ce que c'est pour un trait que d'être génétique. Peut-être pouvons-nous plutôt dire que dans le cas des traits génétiques, les facteurs génétiques sont en quelque sorte plus fondamentaux ou plus importants, même si les facteurs environnementaux sont également pertinents. Après tout, dans le discours ordinaire, nous parlons souvent de « la cause » de quelque chose, sans bien évidemment vouloir explicitement nier l'existence d'autres facteurs pertinents. » (Gifford, 1990, p. 330)

> « […] Il y a des facteurs génétiques et des facteurs non génétiques qui sont causalement pertinents pour n'importe quel trait, un fait reconnu par à peu près tous les commentateurs du concept de maladie génétique […]. Donc ce qui pose problème pour décider si une maladie est génétique ou pas, c'est de savoir si les facteurs causaux qui sont génétiques sont les causes les plus importantes. C'est ce que j'appelle « le problème de la sélection ». Comment

> décider si ce sont les facteurs génétiques ou les facteurs environnementaux qui sont les plus importants dans la production de maladies différentes ? » (Magnus, 2004, p. 234)

Affirmer qu'une maladie est génétique reviendrait à sélectionner le ou les gènes comme le facteur causal le plus important dans l'explication de la maladie, aux dépens d'autres facteurs comme les facteurs environnementaux par exemple. Pour affirmer qu'une maladie est génétique, il faut donc trouver des arguments légitimes pour résoudre le problème de la sélection causale et favoriser les gènes aux dépens de l'environnement. Une telle formulation du problème de la définition des maladies génétiques a deux conséquences sur la suite de la stratégie argumentative : d'une part, la généticisation est rejetée au motif qu'elle constitue une extension illégitime de la définition des maladies génétiques, d'autre part les commentateurs cherchent à déterminer sur quels critères il est légitime de résoudre le problème de la sélection causale en faveur des gènes, c'est-à-dire sur quels critères il est acceptable de définir et de classer des maladies comme « génétiques », en insistant sur l'importance causale du ou des gène(s) au détriment des facteurs environnementaux.

2.1.2. Une stratégie commune : le rejet de la généticisation comme génocentrisme

Dans un premier temps, la généticisation (« toute maladie peut être dite génétique ») est considérée à l'aune du problème de la sélection causale comme l'affirmation que pour toute maladie, le gène est la cause la plus importante dans l'explication de la maladie. Cette affirmation est alors dénoncée comme l'extension d'un génocentrisme dépassé et insupportable. Le génocentrisme est souvent confondu (à tort) avec le déterminisme génétique ou avec le réductionnisme génétique, deux positions philosophiques pour lesquelles ils existent autant de définitions que de nuances.

Nous reprenons ici quelques distinctions fort utiles proposées par Jean Gayon (Gayon, 2009a) entre *un* déterminisme génétique, *le* déterminisme génétique et *le* réductionnisme génétique. On parle d'*un* déterminisme génétique lorsque l'on dit qu'un caractère (par exemple la couleur des yeux ou le sexe) est génétiquement

déterminé : le gène est alors considéré comme une condition nécessaire pour prédire l'apparition du caractère. *Le* déterminisme génétique est une doctrine qui affirme que le phénotype (l'ensemble des caractères apparents) d'un organisme peut être prédit à partir de son génotype. En reprenant les célèbres analyses de Sahotra Sarkar (Sarkar, 1998), Jean Gayon analyse la doctrine du déterminisme génétique dans le contexte de la génétique classique pour distinguer au moins quatre versions différentes du déterminisme génétique :

(1) pour tout locus, deux individus ayant les mêmes allèles à ce locus auront toujours un phénotype identique ;

(2) pour certains locus, deux individus ayant les mêmes allèles à ce locus auront toujours un phénotype identique ;

(3) deux individus identiques pour tous leurs allèles à tous leurs locus (par ex. deux jumeaux vrais) auront toujours un phénotype identique pour tous leurs caractères ;

(4) deux individus identiques pour tous leurs allèles à tous leurs locus (par exemple deux jumeaux vrais) auront toujours un phénotype identique pour certains de leurs caractères.

Les versions (3) et (4) sont des versions moins exigeantes que (1) et (2) puisqu'elles admettent la possibilité qu'un caractère soit déterminé par plusieurs gènes, contrairement à (1) et (2) qui affirment qu'un caractère est entièrement déterminé par un seul locus. Cependant, comme nous l'avons montré dans le chapitre 1 en insistant sur la découverte des phénomènes d'hétérogénéité allélique, d'hétérogénéité génétique, d'hétérogénéité phénotypique, de pléiotropie, et de gènes modificateurs et comme l'affirme Jean Gayon, on peut aisément montrer que toutes ces versions sont fausses :

> « Il est facile de montrer que toutes ces versions du déterminisme génétique sont fausses. (1) est faux car il existe dans tous les organismes de très nombreux caractères dont la genèse est contrôlée non par un unique gène, mais par un complexe de gènes. (2) est faux car on peut toujours imaginer un contexte génétique qui empêche la manifestation d'un caractère monogénique. (3) est faux car il existe bon nombre de caractères qui, chez les jumeaux vrais, dépendent d'une interaction particulière entre gènes et

> environnement. L'énoncé (4), qui est la version la plus faible du déterminisme génétique, est faux pour la même raison : on peut toujours imaginer (et manipuler) une variable environnementale qui empêche l'expression standard d'un génotype donné (par exemple inhiber le développement des yeux d'un animal par une manipulation expérimentale). Cette objection est triviale mais décisive : dans toutes ses formulations, le déterminisme génétique présuppose toujours des conditions environnementales standard. Il serait vain de penser qu'on peut l'amender en intégrant des variables environnementales : c'est précisément la question qui est en jeu. Le déterminisme génétique consiste à soutenir que l'état futur d'un organisme est prédictible sur la base de sa composition génétique. C'est impossible en l'absence de spécification d'un environnement. » (Gayon, 2009a, p. 116–117)

Le déterminisme génétique revient donc par définition à nier l'importance des facteurs environnementaux dans la prédiction du phénotype. Cette doctrine doit cependant être distinguée du réductionnisme génétique, qui se mêle moins de la prédiction que de l'explication.

> « Le réductionnisme génétique est une attitude moins exigeante. Il ne consiste pas à dire que les gènes déterminent entièrement la genèse des traits des organismes, mais que l'explication de ceux-ci doit faire une place essentielle aux facteurs génétiques. Position moins forte, le réductionnisme génétique consiste à dire que la meilleure explication d'un trait biologique est celle qui explicite la manière dont les gènes déterminent ce trait dans un contexte organismique et environnemental donné. Notons au passage que la réduction en question n'est pas la « réduction théorique » chère à certains philosophes des sciences. Il ne s'agit pas de réduire, par exemple, la « théorie » du système immunitaire à la « théorie » génétique. Le réductionnisme génétique est plutôt un pari heuristique consistant à dire qu'un phénomène biologique n'est correctement expliqué qu'à condition de spécifier le rôle que jouent les gènes dans la genèse du caractère en question. » (Gayon, 2009a, p. 117–118)

Deux points de cette citation méritent tout particulièrement d'être soulignés. D'une part, le réductionnisme génétique est une position moins forte que celle du déterminisme génétique : c'est une position presque « méthodologique » qui donne

une importance centrale aux gènes en termes d'explication des traits biologiques et non en termes de prédiction. D'autre part, le terme de « réduction » doit bien être distingué de la « réduction théorique », qui a été introduite par Ernest Nagel dans son ouvrage canonique (Nagel, 1961) et qui envisage la possibilité et les conditions pour réduire des théories de niveau supérieur à des théories de niveau inférieur.

À la lumière de cette distinction, il semble que le concept de génocentrisme soit plus proche du réductionnisme génétique que du déterminisme génétique. Le génocentrisme insiste en effet sur l'importance des gènes en tant que facteur causal majeur de l'explication de tout trait biologique, que ce soit dans le contexte de l'explication du développement d'un organisme, de l'évolution d'une communauté d'organismes à travers le temps ou de l'apparition de maladies. Ainsi, Kelly Smith définit le génocentrisme en ces termes :

> «Essentiellement, le génocentrisme est l'idée que les gènes sont les facteurs explicatifs les plus importants en biologie du fait de leur pouvoir causal unique – que ce soit dans le contrôle de l'ontogénie individuelle (le développement) ou pour rendre compte du fonctionnement anormal des adultes (la maladie) ou encore pour expliquer les changements dans des populations à travers le temps (l'évolution). Cette métaphore tacite à propos du rôle des gènes dans les organismes biologiques explique en grande partie l'enthousiasme aveugle pour les locutions « les gènes de » qui sont employées communément à la fois dans la presse professionnelle et dans le public. » (Smith, 2007, p. 84)

Les critiques adressées au génocentrisme sont généralement moins violentes que celles adressées au déterminisme génétique. Toutefois, lorsque cette notion est appliquée à la santé et à la maladie, il est généralement reproché au génocentrisme ses conséquences sociales. Le génocentrisme aurait ainsi un impact sur l'orientation de la recherche médicale qui privilégierait la recherche de thérapeutiques qui ciblent les gènes au détriment d'autres voies (Magnus, 2004). Il aurait également un impact sur la façon dont les individus vivent leur maladie et se perçoivent comme intrinsèquement malades ou porteurs d'un défaut essentiel (Hull, 1997).

On comprend à présent la démarche argumentative des auteurs qui s'intéressent au concept de maladie génétique et à la généticisation. À partir du

moment où labéliser une maladie comme génétique consiste à résoudre le problème de la sélection causale en faveur des gènes et à affirmer l'importance centrale des gènes dans l'explication de cette maladie, la généticisation ne peut être comprise que comme l'affirmation que les gènes seraient la cause la plus importante dans l'explication de toutes les maladies. Une telle affirmation est considérée comme scientifiquement fausse et semble s'apparenter au génocentrisme dont les conséquences sociales sont particulièrement critiquées. La seule solution consiste donc à rejeter la généticisation au motif qu'elle est une extension pernicieuse du génocentrisme.

2.1.3. Une stratégie commune : restreindre le concept de maladie génétique

L'abandon de la généticisation n'a pas pour autant pour conséquence l'abandon du concept de maladie génétique. Si la plupart des commentateurs du concept de maladie génétique réfutent la généticisation, c'est-à-dire l'extension du concept de maladie génétique à toutes les maladies, ils ne nient pas la légitimité du problème de la sélection causale et du concept de maladie génétique. Ils concluent seulement qu'il faut trouver un critère qui permette de justifier la résolution du problème de la sélection causale en faveur des gènes pour certaines (et non pas toutes) maladies. Il s'agit donc de chercher un sens plus restreint au concept de maladie génétique, c'est-à-dire de délimiter les maladies pour lesquelles on peut affirmer que le gène est la cause la plus importante dans l'explication de la maladie. Ainsi, Kelly Smith s'oppose d'abord à la généticisation des maladies, au motif qu'elle ne repose que sur l'affirmation que les gènes sont un facteur important dans l'explication des maladies, sans expliciter pour quelles raisons les gènes seraient des facteurs plus importants que les facteurs environnementaux.

> « De plus, l'implication des gènes dans tous les traits biologiques est à ce point insidieuse qu'il semble difficile d'envisager un trait dans lesquels les gènes ne sont pas impliqués. Avec une telle conception, il s'ensuit que tout trait est labélisé comme génétique – une position que certains chercheurs médicaux ont adopté vis-à-vis de la maladie. Il est cependant difficile de voir comment une telle description peut apporter une information utile, précisément parce que le trait en question est universel. Tout ce qui est établi, c'est que les gènes

> sont des facteurs essentiels dans n'importe quel processus biologique, un fait que personne ne contestera. Cela n'établit pas que les gènes soient de quelque façon que ce soit *plus* importants que les autres facteurs causaux, ce que semble pourtant impliqué le label « génétique ». Pour être tout à fait franc, si on se fonde sur la seule implication causale, nous avons exactement la même légitimité à décrire tous les traits biologiques come « fondés sur les protéines » ou comme « organique ». Ces descriptions semblent triviales précisément parce qu'elles n'apportent aucune information nouvelle. Ainsi, si nous sélectionnons les gènes comme la meilleure explication causale pour un trait, il faut que ce soit sur une caractéristique unique de leur implication causale et non pas sur le *simple fait* qu'ils soient causalement impliqués. » (Smith, 2007, p. 88–89)

L'essentiel est donc de parvenir à trouver un critère légitime de sélection causale en faveur des gènes pour donner une définition plus restreinte mais adéquate du concept de maladie génétique :

> « La tentation d'adopter une attitude défaitiste vis-à-vis de processus causaux aussi complexes est forte. Cependant il y a aussi un objectif pratique légitime pour de tels « raccourcis explicatifs » – ne serait-ce que le fait que nous ne pouvons pas décrire la situation causale de façon exhaustive à chaque fois que nous souhaitons affirmer quelque chose au sujet de la ou des cause(s) de la maladie. Par ailleurs, ces raccourcis sont relativement inoffensifs du moment que l'on garde précisément à l'esprit quelles informations ils apportent ou pas au sujet du système causal en question. Les problèmes peuvent cependant surgir assez facilement, dès lors que l'on utilise un critère vague ou simpliste pour la sélection causale ou lorsque le critère utilisé est mal interprété par le public attendu. Il est donc crucial que la sélection causale soit faite sur la base d'un critère explicite qui aura été évalué de façon critique. Si nous identifions une maladie comme génétique, nous devrions être capable de donner une description claire de ce que cela implique précisément et de pourquoi cette caractérisation nous fait gagner davantage en transmettant de façon concise une information causale importante qu'elle ne nous fait perdre en étant incomplète. » (Smith, 2007, p. 84)

Nous allons à présent passer en revue les différentes stratégies proposées pour résoudre le problème de la sélection causale en faveur des gènes dans les explications de la maladie et pour ainsi renouveler le sens du concept de maladie génétique.

2.2. Trois approches pour résoudre le problème de la sélection causale

Si une même stratégie philosophique réduit la question de la généticisation et du concept de maladie génétique à un problème de sélection causale, les approches diffèrent pour justifier la résolution du problème de la sélection causale en faveur des gènes. Trois stratégies peuvent être identifiées. La première stratégie consiste à établir une relation entre le gène et la maladie (nécessité, suffisance, etc.) qui permette de sélectionner le gène comme la cause la plus importante de la maladie et de reléguer l'environnement au rang de simple condition. Cette première stratégie que nous appelons l'approche cause/condition aboutit donc à un concept de maladie génétique en soi : dans l'explication de certaines maladies, cette relation entre le gène et la maladie existe en vertu de quoi on peut légitimement les qualifier de « maladies génétiques ». Il y a donc des maladies génétiques et des maladies non génétiques. Par opposition, la seconde stratégie ne cherche pas à expliquer l'apparition de la maladie comme un effet de la présence d'une anomalie génétique, mais à expliquer la différence entre des individus malades et des individus sains par des différences génétiques – et c'est pourquoi nous l'appelons « différentialiste ». Cette seconde stratégie, qui attribue des différences phénotypiques entre individus aux différences génétiques est donc toujours relative au choix d'une population de contraste. Elle n'aboutit pas à une délimitation en soi entre maladie génétique et maladie non génétique : une même maladie peut être considérée comme génétique ou non génétique, en fonction du choix de la population de contraste. La troisième stratégie est purement pragmatique : elle reconnaît l'extrême dépendance au contexte des explications génétiques et considère la relativité de ces explications à la lumière de cette dépendance contextuelle. Nous affirmons cependant qu'aucune de ces approches ne permet de définir de façon satisfaisante le concept de maladie génétique. Pour prouver cet échec, nous avons choisi un représentant philosophique de chaque approche dont nous allons examiner en détail la démarche.

2.2.1. L'approche cause/condition : définir la maladie génétique en soi

La première stratégie que nous appelons l'approche cause/condition consiste donc à trouver un critère qui permette de considérer que le gène est la cause la plus importante de la maladie et de reléguer l'environnement au rang de simple condition. L'approche d'Henrik Wulff exemplifie parfaitement cette stratégie. Il distingue ainsi cinq types de maladies (Wulff, 1984) :

(1) « Une anomalie génétique est un déterminant nécessaire et suffisant de la maladie, indépendamment de son environnement » : l'exemple donné par Wulff pour ce cas est le syndrome de Klinefelter, une aneuploïdie qui se caractérise par la présence d'un chromosome X supplémentaire chez des personnes de sexe masculin (possédant par un ailleurs un chromosome X et un chromosome Y) et qui peut entraîner une infertilité.

(2) « Une anomalie génétique est un déterminant nécessaire mais non suffisant de la maladie et les facteurs environnementaux qui complètent la matrice causale se produisent normalement » : un exemple serait la phénylcétonurie, une anomalie du gène PAH entrainant une mutation de la phénylalanine hydroxylase – mutation dont les effets ne s'expriment que parce que nos régimes alimentaires comportent majoritairement de la phénylalanine.

(3) « Une anomalie génétique est un déterminant nécessaire mais non suffisant de la maladie et les facteurs environnementaux qui complètent la matrice causale se produisent rarement » : l'exemple donné ici est l'hypersensibilité au suxaméthonium (un produit anesthésique) responsable d'accidents graves à l'induction de l'anesthésie. L'anomalie génétique est ici un déterminant nécessaire

mais non suffisant puisqu'il faut être exposé au suxaméthonium pour développer la maladie. Cependant l'exposition au suxaméthonium est elle-même très rare.

(4) « Une anomalie génétique est un facteur non nécessaire et non suffisant de la maladie (une INUS)». Nous rappelons ici rapidement le concept de facteur INUS élaboré par Mackie (Mackie, 1965) : un facteur INUS est la part insuffisante mais nécessaire d'une condition qui est elle-même non nécessaire mais suffisante pour garantir la production d'une effet, en l'occurrence de la maladie. Wulff donne ici l'exemple de la spondylarthrite ankylosante. Dans cette maladie auto-immune et rhumatologique à l'étiologie mal connue, on sait que l'antigène HLA57 augmente considérablement la probabilité de développer la maladie (90% des malades sont ainsi porteurs de l'antigène HLA57, ce dernier étant présent dans moins de 10% de la population générale). Cependant il existe des individus qui sont porteurs de l'antigène HLA57 et qui ne développent pas la maladie et il y a des individus qui sont malades et qui n'ont pas l'antigène HLA57.

(5) « la maladie n'est pas du tout génétiquement déterminée, ce qui veut dire que les facteurs environnementaux sont nécessaires et suffisants ». C'est l'exemple de l'empoisonnement à l'arsenic qui est sollicité pour illustrer ce dernier cas de figure. Il y a certainement des mutations génétiques qui doivent moduler la réponse immunitaire à l'arsenic, agent environnemental. Mais il suffit d'avoir été contaminé par l'arsenic pour développer la maladie et seul l'arsenic entraine le développement de cette maladie.

Selon Wulff, à partir de ces différentes situations, on peut donc distinguer un sens fort de la maladie génétique et un sens plus faible :

> « J'espère avoir montré par ces exemples que le concept de maladie génétique est un concept raisonnable. Selon la définition forte, une maladie est génétique, si l'anomalie génétique est à la fois un facteur causal nécessaire et suffisant, ce qui signifie que la maladie se développera dans n'importe quel environnement. La condition minimale est que le facteur génétique soit un déterminant dans tous les patients qui ont la maladie. Les autres maladies sont partiellement génétiquement déterminées puisque l'anomalie génétique est un facteur INUS. Dans ces cas, l'anomalie génétique est seulement une partie de la matrice causale chez certains des patients qui ont la maladie. » (Wulff, 1984, p. 197)

Wulff utilise ici le critère de la nécessité et de la suffisance pour distinguer selon les cas qui du gène ou de l'environnement a le rôle de la cause ou de la condition. Dans les cas (1), (2) et (3), l'anomalie génétique est considérée comme la cause de la maladie et l'environnement comme une des conditions de la matrice causale. En revanche, dans les cas (4) et (5), c'est l'environnement qui est considéré comme une cause nécessaire et/ou suffisante, tandis que l'anomalie génétique est une simple condition ou n'entre même pas dans la matrice causale (cas 5). Seul le cas (1) où l'anomalie génétique est à la fois nécessaire et suffisante pour développer la maladie répond à la définition forte du concept de maladie génétique, les cas (2) et (3) sont néanmoins considérés comme tels parce que l'anomalie génétique est nécessaire pour développer la maladie : on ne peut pas développer la maladie sans être porteur de l'anomalie génétique.

L'argumentation de Wulff est aisée à balayer au regard de nos connaissances scientifiques actuelles. Le cas (1) semble triplement contestable. Le syndrome de Klinefelter n'est d'abord pas toujours considéré comme une maladie : l'infertilité n'est pas systématique et en dehors de l'infertilité, les autres symptômes de la maladie (retard pubertaire, faible pilosité, petite taille, possible gynécomastie, petite taille des testicules) ne sont pas nécessairement visibles ou handicapants dans la vie quotidienne, ce qui explique que la plupart des adultes ne découvrent que très tard être porteurs du syndrome de Klinefelter ou lorsqu'ils ont un projet parental. Nonobstant cette discussion, le syndrome de Klinefelter n'est pas à proprement parler une anomalie génétique : c'est une aneuploïdie, c'est-à-dire une anomalie du nombre

de chromosomes (un chromosome étant par définition la somme de plusieurs dizaines de gènes) due à une mauvaise ségrégation des chromosomes lors de la méiose. Enfin, il y a plusieurs formes de la maladie : la forme majoritaire donne le caryotype 47, XXY dans toutes les cellules, mais 20% des patients atteints d'un syndrome de Klinefelter ont une forme mosaïque. Dans ce cas, l'anomalie est présente dans certaines cellules qui ont le caryotype 47, XXY alors que les autres cellules sont normales (caryotype 46,XY), car l'anomalie s'est produite après la formation du zygote. Les symptômes sont souvent atténués dans la forme mosaïque mais même dans la forme majoritaire, les patients ne sont pas tous infertiles. La relation de nécessité est donc fortement discutable ici. Dans les cas (2) et (3), nous avons déjà expliqué plus haut que certaines anomalies du gène PAH sont sans conséquence sur le fonctionnement de la phénylalanine hydroxylase et qu'il était possible d'avoir un phénotype d'hyperphénylalanisme sans mutation du gène PAH. De façon plus générale, nous avons insisté dans le premier chapitre de cet exposé sur la remise en question du modèle de la maladie monogénique dont s'inspire la définition de Wulff.

Indépendamment de cette analyse des exemples proposés ici par Wulff, la question de la nécessité soulève d'autres problèmes plus fondamentaux. Comme le montre Gifford (Gifford, 1990), quand bien même il existerait une relation de nécessité entre une anomalie génétique et une maladie, d'autres facteurs seront probablement nécessaires à l'apparition de la maladie comme l'absence d'une température de 1000 degrés par exemple. L'absence d'une température de 1000 degrés peut certes être considérée comme une condition universellement nécessaire au développement de la vie et être rejetée en vertu de ce principe dans la catégorie des conditions *ceteris paribus*, comme semble le faire Wulff. Mais même en excluant les conditions nécessaires universelles, on peut imaginer aisément d'autres conditions nécessaires au développement de la maladie. Ainsi, Kelly Smith donne l'exemple de la concentration précise d'une protéine ou de sa distribution qui peut être un facteur causal nécessaire à l'apparition d'un trait (Smith, 2007). On peut mettre cet exemple de la concentration précise d'une protéine en lien avec la notion d'effet-seuil que nous avons discutée dans l'exemple de la phénylcétonurie ou de l'anémie falciforme : dans l'anémie falciforme, il faut atteindre un certain seuil de concentration d'une protéine mutée à partir de laquelle la maladie se déclenche. Atteindre ce seuil est une condition

nécessaire qui ne dépend pas uniquement de la présence d'une mutation dans le gène responsable de la production de la protéine. Le critère de la nécessité peut donc être considéré comme un échec : d'une part, la littérature scientifique contemporaine questionne très fortement l'existence d'exemples où une telle relation de nécessité entre gène et maladie existe (Dipple et McCabe, 2000b) et d'autre part, parviendrait-on à produire de tels exemples, qu'il faudrait encore justifier en quoi cette cause génétique nécessaire serait plus importante que d'autres causes nécessaires à l'apparition de la maladie.

D'autres critères que la nécessité ont été proposés, dont Kelly Smith a fait une liste quasi-exhaustive (Smith, 2007). Nous renvoyons à la lecture de Kelly Smith pour une explicitation exhaustive de tous ces critères et la discussion de leur insuffisance. Nous allons nous arrêter ici seulement sur la discussion de deux critères qui reviennent fréquemment dans les argumentaires des philosophes : les critères de suffisance et de manipulabilité.

(1) Le gène est la cause suffisante de la maladie. Dans le chapitre 1, nous avons montré que des phénomènes comme la pénétrance variable que l'on retrouve aussi bien dans la maladie de Huntington que dans le cas de la phénylcétonurie ou dans celui de l'anémie falciforme pour rappeler quelques exemples connus, explique que la présence de l'anomalie génétique suffise rarement à provoquer la maladie. On trouve ainsi pour chacune de ces maladies des patients qui ont l'anomalie génétique, mais qui ne développent pas la maladie. Par ailleurs, la question de la suffisance pose clairement problème dans des cas comme la phénylcétonurie ou la sensibilité au sumaxéthonium, puisque dans ce type de cas, la maladie ne se déclenche que si l'individu porteur de l'anomalie génétique rencontre un environnement défavorable (ingestion de phénylalanine, rencontre avec un produit anesthésique). Il n'y a pas de raison dans ces exemples de considérer que le gène ou que l'environnement soit une cause suffisante, à moins d'introduire la notion de population de contraste et de fixer l'environnement comme une condition *ceteris paribus*. Dans le cas de la phénylcétonurie, cela revient à dire que le gène est la cause suffisante de la maladie dans un environnement où les individus sont tous soumis à un régime qui contient de la phénylalanine. Le critère de la suffisance pose donc des problèmes d'application tels

qu'ils nécessiteraient pour être réglés d'introduire la notion de population de contraste – une notion qui pose elle-même des difficultés spécifiques que nous discuterons plus loin, lorsque nous expliciterons le deuxième grand type de stratégie (la stratégie différentialiste) qui consiste justement à définir le concept de maladie génétique par rapport à une population donnée.

(2) Le gène est la cause la plus manipulable de la maladie. Le concept de manipulabilité revient fréquemment dans la littérature sur le concept de maladie génétique (Caplan, 1992 ; Magnus, 2004) – probablement en raison de son attrait pratique. Parce que la médecine est d'abord intéressée par les possibilités thérapeutiques, il semble que définir la cause la plus importante comme la cause la plus manipulable garantisse l'identification de nouvelles cibles thérapeutiques. En l'occurrence, le critère de manipulabilité génétique est donc lié au concept de thérapie génique qui suppose de pouvoir remplacer un gène défaillant par un gène non muté. C'est là que commencent les difficultés. Parle-t-on de manipulabilité pratique ou de manipulabilité théorique ? Si on parle de manipulabilité pratique, les avancées techniques actuelles sont certes encourageantes pour la thérapie génique, mais encore au stade des essais cliniques. Les premiers cas de thérapie génique ont eu lieu en 1999 avec les expérimentations du professeur Alain Fischer sur des enfants atteints d'immunodéficience sévère combinée, une maladie du système immunitaire lié à des mutations du gène ILR2G sur le chromosome X, qui entraîne un défaut de production de deux grandes classes de cellules immunitaires (les lymphocytes T et les cellules NK), aboutissant à une défaillance majeure du système immunitaire. Les enfants qui en sont atteints sont aussi appelés « enfants bulles » car ils sont placés dès leur naissance dans des chambres stériles, en attentant d'être traités par greffe de moelle osseuse allogénique. Le traitement par greffe de moelle osseuse est cependant un processus compliqué. Entre autres il est parfois difficile de trouver un donneur avec un système HLA compatible dans les familles peu nombreuses. Il faut alors se tourner vers une greffe de moelle osseuse dont l'histocompatibilité est imparfaite, ce qui augmente les risques du traitement : la reconstitution du système immunitaire est souvent imparfaite, le risque de la maladie du greffon contre l'hôte est aussi plus important. C'est pourquoi en 1999, l'équipe d'Alain Fisher à l'hôpital Necker a expérimenté un

traitement par thérapie génique sur neuf enfants pour lesquels aucun donneur de moelle osseuse histocompatible n'avait été trouvé. La thérapie génique proposée se décompose en plusieurs étapes : il faut d'abord *ex vivo* insérer une copie du gène normal dans les précurseurs médullaires de ces enfants à l'aide d'un vecteur rétroviral avant de réinjecter ces cellules dans la moelle des enfants. La thérapie génique a permis de restaurer le système immunitaire de tous les enfants traités, mais quatre ont développé par la suite une leucémie aiguë. Sur ces quatre enfants, trois sont aujourd'hui guéris tandis que le quatrième est décédé. Si l'équipe d'Alain Fischer a pu ainsi publier à juste titre que huit patients sur neuf en 2010 étaient toujours guéris (Hacein-Bey-Abina *et al.,* 2010), le traitement entraîne un important risque de développer une leucémie aiguë, du fait d'une mauvaise insertion du vecteur dont le promoteur aurait été inséré devant un gène oncogène. Si la thérapie génique est donc bien un succès, c'est un succès limité : même dans le cadre de l'immunodéficience sévère combinée, le traitement par thérapie génique n'est pas conseillé comme traitement de première intention, mais seulement lorsque la greffe de moelle osseuse allogénique histocompatible est impossible. Si on revient à la question de la manipulabilité pratique, on peut donc affirmer que la thérapie génique est loin d'être généralisée à toutes les maladies génétiques. Si on adopte le point de vue de la manipulabilité théorique ou en principe, alors effectivement le gène apparaît comme une cause manipulable de la maladie. Rien ne permet en revanche d'affirmer que ce sera la cause la plus manipulable de la maladie. Si on revient à l'exemple de la phénylcétonurie, il semble que même en admettant que la thérapie génique soit possible ou concevable dans le cadre de la phénylcétonurie, un simple traitement par régime exempt de phénylalanine semble beaucoup moins coûteux, dangereux et plus efficace. Le critère de la manipulabilité semble donc aussi échouer à résoudre le problème de la sélection causale en fonction des gènes et à redéfinir le concept de maladie génétique. Nous allons à présent nous tourner vers la seconde stratégie, qui ne cherche pas à définir un concept de maladie génétique en soi, mais se place d'emblée dans un cadre où le concept de maladie génétique ne peut être que relatif à une population de contraste donnée.

2.2.2. L'approche différentialiste : définir la maladie génétique par rapport à une population de contraste

La seconde stratégie que nous appelons « explication différentialiste » (aussi appelée l'explication en termes de « cause qui fait la différence » ou *« difference-maker explanation »)*, consiste à définir un concept de maladie génétique relatif à une population donnée, c'est-à-dire à trouver des cas pour lesquels les différences entre les individus malades et les individus sains sont causées par le gène (Gifford, 1990 ; Smith, 2001, 2007 ; Waters, 2007). Gifford appelle ce critère « le critère du facteur de différence » (*« Difference factor criterion »)* : « un trait est génétique (respectivement à la population P) si ce sont des facteurs génétiques « qui font la différence » entre les individus qui ont le trait et le reste de la population P ». Waters en a donné une version plus sophistiquée mais Gifford comme Waters se sont principalement intéressés à la question des « traits génétiques » plutôt qu'au concept des maladies génétiques. C'est pourquoi nous présenterons et discuterons ici l'approche de Kelly Smith qu'il appelle « approche épidémiologique » du concept de maladie génétique.

Kelly Smith part du postulat suivant : pour être satisfaisante, la description du concept de maladie génétique doit exemplifier deux intuitions que nous posséderions au sujet du concept de maladie génétique.

(1) L'intuition individuelle : si une maladie est génétique, cela doit signifier que ceux qui ont le gène sont davantage susceptibles de développer la maladie que ceux qui ne l'ont pas.

(2) L'intuition au niveau de la population : si une maladie est génétique, cela doit signifier que dans la population des individus malades, une majorité de cas implique causalement le gène.

À partir de ces deux intuitions, Kelly Smith utilise un exemple fictif de maladie génétique (voir Tableau 5). Il est intéressant de remarquer que l'exemple choisi par Kelly Smith est un exemple de maladie génétique relativement complexe, en tout cas, assez éloigné de l'exemple de la maladie monogénique simple, puisque dans cet exemple a) tous ceux qui ont le gène (1000 personnes) ne développent pas la maladie (phénomène de la pénétrance incomplète et des porteurs asymptomatiques), b) certains ont le gène mais développent la maladie pour d'autres raisons (hétérogénéité

génétique), c) et enfin certains sont malades et n'ont pas le gène (pénétrance incomplète).

	Malade	Non Malade	Total
Gene présent, causalement impliqué	760	0	760
Gène présent, non impliqué	40	200	240
Gène absent	450	8 550	9 000
Total	1250	8 750	10 000

Tableau 5 : Présentation d'un exemple fictif de maladie génétique selon Kelly Smith (Source : Kelly Smith, 2007)

À partir de ce tableau, Kelly Smith rappelle deux notions d'épidémiologie régulièrement utilisées :

- *la Fraction Etiologique dans la Population*: c'est le pourcentage d'individus qui ont la maladie et qui vont avoir la maladie à cause du gène. En l'occurrence dans le tableau ci-dessus, la fraction étiologique de la population correspond à la division suivante : nombre d'individus qui ont la maladie et où le gène est causalement impliqué / nombre total d'individus malades, soit 760/1250 = 61%
- *le Risque Attribuable :* c'est le pourcentage d'individus qui ont le gène et qui vont avoir la maladie à cause du gène. En l'occurrence, dans le tableau ci-dessus, le risque attribuable se calcule ainsi : nombre d'individus qui ont la maladie et qui ont le gène / nombre d'individus qui ont le gène : 760/1 000 = 76%

Que signifient ces indicateurs et la valeur de ces pourcentages ? Avoir une fraction étiologique de la population supérieure à 50% signifie que la plupart des individus malades ont la maladie à cause du gène ; avoir un risque attribuable supérieur à 50% signifie que la plupart des individus qui ont le gène ont la maladie à cause du gène.

Pour être certain d'identifier une maladie génétique et répondre à nos intuitions sur ce concept, il faut donc une maladie qui remplisse deux critères à la fois :

avoir une fraction étiologique de la population et un risque attribuable tous les deux élevés. Reste à savoir comment déterminer le seuil nécessaire à atteindre pour pouvoir parler de maladie génétique : 50% ? 70% ? 100% ? Kelly Smith va alors proposer deux définitions-seuils permettant de définir une forme de « continuum » de la maladie génétique.

Le premier seuil correspond à ce que Kelly Smith appelle une définition « Pratiquement Suffisante » de la maladie génétique : une maladie est génétique si le risque attribuable du gène et la fraction étiologique de population sont de 100%. Dans ce cas de figure, cela signifie que tous les individus avec le gène ont la maladie à cause de l'implication du gène et qu'aucun des individus malades n'a la maladie pour une autre raison que le gène. Ce cas de figure, et c'est pourquoi Kelly Smith l'appelle « description pratiquement suffisante » est très proche de la définition proposée par Wulff, où il faut que l'anomalie génétique soit à la fois nécessaire et suffisante pour développer la maladie. Cela ne revient cependant pas à dire que les gènes sont nécessaires ou suffisants pour développer la maladie, il suffit que les gènes soient nécessaires et suffisants dans une population précise. Les critères de nécessité et de suffisance sont donc relatifs à une population donnée : autrement dit, il est tout à fait possible pour une même maladie d'être génétique dans un contexte donné et de ne pas l'être dans un autre. Kelly Smith reconnaît cependant qu'avec une définition aussi stricte, sa description a peu de chances de pouvoir s'appliquer, même dans des cas où il semble que nous ayons d'excellentes raisons de parler de maladie génétique. Il propose alors de mettre en place un deuxième seuil et donne une description « Minimalement Épidémiologique » de la maladie génétique : une maladie est génétique quand la fraction étiologique dans la population et le risque attribuable sont supérieurs à 50%. Dans ce cas, la majorité des individus qui ont le gène ont la maladie à cause de l'implication du gène et la majorité des individus malades ont la maladie à cause de l'implication causale du gène.

Kelly Smith décrit ainsi le continuum génétique rendu possible par la mise en place de ces deux valeurs seuils. À l'une des extrémités du continuum, on trouve les maladies qui ne sont pas génétiques, c'est-à-dire qui ne satisfont aucun des deux seuils : Kelly Smith donne ici l'exemple de l'obésité. Selon lui, il y a de fortes chances pour que dans nos sociétés occidentales, l'obésité ne puisse pas se qualifier comme

maladie génétique, bien qu'il soit impossible de le prouver faute de données sur le nombre d'individus sains possédant les gènes associés à l'obésité. En effet, il est probable que la majorité des individus malades le soient non pas à cause de l'implication du gène, mais bien à cause de leur régime alimentaire. À l'autre extrémité du continuum, on trouve des maladies génétiques à la fois au sens « minimalement épidémiologique » (ME) et au sens « pratiquement suffisant » (PS), comme le syndrome de Klinefelter. Au milieu de ce continuum, on trouve enfin des maladies comme la mucoviscidose. Il y a plus de mille mutations du gène CFTR (responsable de la production du canal chlore) impliquées dans la mucoviscidose, sans que l'on parvienne à déterminer (sauf pour de rares exceptions) de bonnes corrélations statistiques entre le profil des mutations et la sévérité des manifestations phénotypiques de la maladie. Par ailleurs, certaines personnes dont le génotype est muté ne développent pas la maladie. Bien que par manque de données sur la population normale, on ne puisse pas donner de chiffres, on peut supposer que pour une telle maladie, la fraction étiologique dans la population et le risque attribuable ne sont pas de 100% mais sont supérieurs à 50%. La maladie n'est donc pas génétique au sens d'une définition « pratiquement suffisante » mais se qualifie comme « maladie génétique » au sens d'une définition « minimalement épidémiologique ».

Que peut-on reprocher au concept de maladie génétique tel qu'il est formulé chez Kelly Smith ? Son approche, comme toutes les approches différentialistes, permet d'adapter les concepts de nécessité et de suffisance relativement à une population donnée, s'accorde bien à une vue épidémiologique des maladies reposant en partie sur les idées de variance et de covariance si chères aux biologistes, peut être testée empiriquement et abolit toute distinction absolue entre maladies monogéniques et maladies polygéniques, ce qui semble bien s'ajuster à la conception scientifique contemporaine des maladies génétiques. Kelly Smith lui-même souligne cependant deux limites de son approche.

Le premier inconvénient de la position différentialiste concerne l'analyse des cas individuels. En effet, comme le reconnaît Kelly Smith, adopter la position différentialiste suppose qu'on ne peut strictement rien dire des cas individuels. Autrement dit, si un patient insiste pour savoir quelle est la cause de sa maladie, il est impossible de lui répondre sans faire appel aux données de la population générale.

Tout en reconnaissant cette difficulté, Kelly Smith soutient qu'il n'y a pas d'alternative possible : pour sélectionner une cause dans un cas individuel, il faut importer d'une façon ou d'une autre des données générales sur la population. Nous ne discutons pas ce point, c'est la seconde limite qui nous intéresse le plus.

La seconde limite est une conséquence directe d'une approche relative à une population de contraste : il n'y a pas de maladies génétiques en soi. De ce point de vue, nous voulons d'ailleurs remarquer que le terme de continuum nous paraît mal choisi : il semble accréditer l'existence de maladies génétiques en soi, indépendamment de toute considération en termes de population ; alors que c'est uniquement en fonction d'une population que l'on pourra parler de maladie génétique. L'intolérance au lactose par exemple sera une maladie génétique aux États-Unis et une maladie environnementale en Asie. En effet, aux États-Unis, la consommation de lait est fréquente, tandis que la mutation est relativement rare et c'est exactement le contraire en Asie. Ce qui pose problème ici, ce n'est pas le choix de définir une maladie comme génétique, relativement à un contexte donné. Le problème est de savoir sur quels fondements on choisira la population de contraste. Pour éclairer cette difficulté, David Magnus imagine la situation suivante (Magnus, 2004) : supposons que la moitié d'un village présente soudainement des crampes, une diarrhée et une déshydratation conséquente. On envoie sur place une équipe d'experts médicaux et d'épidémiologistes pour déterminer les causes de la maladie. Il est découvert que le château d'eau a été contaminé. Spontanément, il semble évident que la cause de la maladie est le parasite présent dans l'eau contaminé. Mais, si la totalité du village a été exposé au parasite, seulement la moitié du village est tombée malade : en admettant qu'un facteur génétique soit trouvé qui explique la susceptibilité de la moitié du village à l'infection parasitique, c'est bien le gène qui est la cause de la maladie et non le parasite. La conclusion paraît contre-intuitive et pourtant Kelly Smith le confirme : dans un tel cas, les facteurs génétiques apparaissent comme « pratiquement suffisants » dans l'apparition de la maladie, alors que la contamination par l'eau parasitée est un facteur environnemental partagé par tous les membres du village et peut donc être considérée comme une condition *ceteris paribus* de la population de contraste choisie. Tout en reconnaissant la pertinence de l'exemple de David Magnus, Kelly Smith discute le bien fondé de l'exemple, affirmant

que le choix de la population de contraste est mauvais : si on choisit plutôt de comparer la population du village à la population du village voisin dont l'eau n'a pas été contaminée, alors c'est le parasite qui sera la cause de la maladie. David Magnus ne contredit pas ce point, mais souligne ce qui est donc le véritable problème de la stratégie différentialiste : comment choisir la population de contraste ?

> « Une façon de défendre ce type d'analyse serait d'affirmer que nous nous trompons en identifiant la population de l'analyse avec celle des villageois. Si l'on avait choisi comme population la planète entière ou les villages voisins, l'analyse en termes de variance auraient fonctionné […] Cette défense est clairement inadéquate. Le problème de départ de sélection est de décider sur quels fondements on choisit une cause parmi plusieurs comme la plus importante. Au mieux, cette nouvelle approche repousse le problème d'un niveau ; au lieu d'avoir à sélectionner entre des causes, nous avons maintenant à choisir entre des populations ou entre des classes de contraste. Si nous prenons comme population pertinente le village, l'approche statistique définira notre hypothétique maladie comme génétique. Si nous prenons une population d'individus sains qui n'ont pas été exposés au même pathogène comme contraste, alors le pathogène est la cause de la maladie. En effet, il manque quelque chose de plus fondamental dans cette analyse. Dans ce cas, notre hypothétique équipe d'experts médicaux et de scientifiques ne pouvait pas à l'avance savoir quelle population il serait pertinent de choisir pour que l'analyse statistique fonctionne. Vraisemblablement, dans un village où à peu près la moitié des habitants est tombée malade, essayer de chercher la différence entre ceux dans le village qui sont tombés malades et ceux qui ne le sont pas est une approche initiale qu'il serait plausible de choisir. C'est seulement après que l'on a découvert que l'eau contient un pathogène (et que celui-ci a été identifié comme la cause) que le défenseur de l'approche statistique serait capable de faire appel à une population plus large qui sauvera l'analyse. » (Magnus, 2004, p. 235–236)

C'est précisément ce dernier point qui nous paraît fondamental. On peut tout à fait vouloir défendre un concept relatif de maladie génétique, mais alors il faut s'interdire de parler de continuum génétique et de définir des classes de maladie, puisqu'une maladie qui sera génétique dans une population à un moment donné

pourrait très bien ne pas être génétique dans une autre population ou si l'environnement de la population change. Accepter la relativité de cette analyse suppose d'aller jusqu'au bout du raisonnement et d'accepter dans ce cas que le choix de la population de contraste n'a rien d'intuitif – sinon *a posteriori*, comme le souligne l'argumentation de David Magnus. Il faut donc reconnaître que le problème de la sélection causale a été repoussé d'un cran au choix de la population de contraste et que celui-ci ne peut que dépendre d'intérêts relatifs au type et au contexte de la question posée. Autrement dit, accepter l'approche différentialiste mène nécessairement à la position pragmatique que défend Lisa Gannett (Gannett, 1999), et que nous allons discuter dans la suite de cet exposé.

2.2.3. L'approche pragmatique : définir la maladie génétique par rapport à un contexte explicatif

Finalement, la troisième stratégie revient à assumer une approche pragmatique jusqu'au bout, ce qui revient à dire qu'il n'y a pas de fondement ontologique au concept de maladie génétique – si on choisit le gène comme facteur causal le plus important, c'est seulement en fonction d'un contexte explicatif particulier. Cette stratégie repose essentiellement sur la conception des « questions-pourquoi » de Van Fraassen. Selon Van Fraassen (Van Fraassen, 1980), si les théories scientifiques sont indépendantes du contexte et permettent la prédiction, les explications causales sont au contraire complètement dépendantes du contexte. Répondant à la question pourquoi, elles sont construites et orientées en fonction de ces questions et des intérêts de ceux qui la posent. En reprenant les différents types d'explication des maladies génétiques, Lisa Gannett identifie une triple dépendance contextuelle aux explications génétiques (Gannett, 1999) :

(1) Lorsqu'on cherche à expliquer les maladies génétiques en soi, les gènes ne peuvent être considérés comme des causes de la maladie que de façon relative à un faisceau de conditions environnementales et génétiques, considérées comme des conditions *ceteris paribus*.

(2) Lorsqu'on cherche à expliquer les maladies génétiques relativement à une population donnée, c'est-à-dire qu'on cherche à expliquer les différences phénotypiques par des différences génétiques, il semble possible d'échapper au

problème d'une distinction arbitraire entre cause et condition comme c'était le cas en (1). Mais cela revient à reconnaître la dépendance des explications génétiques au choix de la population de contraste.

(3) Ces deux types d'explications (1) et (2) sont elles-mêmes contextuellement dépendantes d'une troisième façon : elles sont épistémiquement relatives à l'affirmation que la maladie peut apparaître dans tous les environnements considérés ; alors même qu'il est toujours possible d'imaginer un environnement alternatif où la maladie n'apparaîtrait pas. Sur une planète différente où la phénylalanine n'existerait pas mais où la vie humaine serait possible, la phénylcétonurie ne pourrait pas exister, pas plus que son explication génétique.

Partant de cette triple relativité des explications génétiques, Lisa Gannett s'intéresse à la question de la généticisation des maladies qu'elle explique comme une conséquence des premiers succès de la biologie moléculaire, des intérêts financiers des laboratoires de génétique, mais aussi d'une conception occidentale de la médecine qui met l'accent sur les facteurs internes de la maladie plutôt que sur les facteurs externes. David Magnus reprend d'ailleurs cette analyse en des termes similaires analysant l'intérêt financier qu'il y a à affirmer pour un laboratoire de recherche qu'il travaille sur des maladies génétiques, ce qui est perçu comme une source potentiel de développements diagnostiques (tests génétiques) et thérapeutiques (thérapie génique). Parler de maladie génétique peut également être dans l'intérêt des États qui, par ce biais, orientent la responsabilité vers l'individu plutôt que vers la santé publique (Magnus, 2004).

L'analyse de Lisa Gannett a certainement le mérite de souligner des processus sociologiques à l'œuvre dans la généticisation, un terme qui a d'ailleurs été forgé dans le contexte de l'analyse sociologique, comme nous le rappelions dans le chapitre 1. Mais si nous allons jusqu'au bout de cette approche pragmatique, on en conclut logiquement que le concept de maladie génétique n'est pas nécessaire.

> « Affirmer, comme je l'ai fait, que les traits sont désignés comme « génétiques » pour des raisons pragmatiques ne revient pas à nier que les gènes soient des agents causalement efficaces […]. Ce que je nie, c'est que des termes comme « trait génétique », « maladie génétique », « les gènes de » soient « objectifs », si l'on entend par « objectifs », dépourvus de contenu

> pragmatique. J'affirme que la façon dont la distinction cause-condition est tracée, dont la population est sélectionnée et dont certaines tendances en recherche sont suivies, sont des choix qui sont influencés par les objectifs, les intérêts et les orientations de ceux qui font ces choix. En appréciant les dimensions pragmatiques des explications génétiques, nous sommes forcés de reconnaître leur contingence et le besoin de s'interroger sur les désirs qui focalisent notre attention sur les causes génétiques ». (Gannett, 1999, p. 370-371)

Pour le dire différemment, l'approche pragmatique offre une solution pratique au problème de la sélection causale mais elle signifie aussi qu'une définition ontologique du concept de maladie génétique n'est plus nécessaire, puisqu'une maladie sera considérée comme génétique seulement de façon relative à un contexte d'explication particulier ou au choix d'une certaine population de contraste.

2.3. Il faut abandonner le concept de maladie génétique

2.3.1. Il n'y a pas de critère pour définir objectivement le concept de maladie génétique

Reprenons à présent notre analyse. Nous avons établi que la plupart des philosophes qui s'intéressent au concept de maladie génétique ont adopté une stratégie argumentative commune, visant à considérer que définir une maladie comme génétique revient à trouver des fondements légitimes pour résoudre le problème de la sélection causale en faveur des gènes. Une fois ce point acquis, ils considèrent la généticisation comme une extension illégitime et génocentriste du concept de maladie génétique qui reviendrait à affirmer que pour toute maladie, les gènes seraient la cause prédominante de la maladie. Le mouvement suivant consiste donc à combattre la généticisation, tout en défendant la possibilité de parvenir à une définition plus stricte et plus restreinte du concept de maladie génétique. Il s'agit donc de chercher des critères qui permettent de résoudre le problème de la sélection causale en faveur des gènes pour un groupe de maladies alors labélisées « maladies génétiques ». Trois stratégies sont dès lors possibles pour résoudre le problème de la

sélection causale. La première stratégie cherche à définir un concept de maladie génétique en soi en distinguant les gènes comme cause de la maladie pour reléguer l'environnement au rang de condition *ceteris paribus*. Mais tous les critères utilisés (nécessité, suffisance, manipulabilité, etc.) échouent à rendre cette distinction non arbitraire. La deuxième stratégie plus prometteuse est différentialiste, elle cherche à expliquer non pas les maladies, mais les différences entre individus sains et individus malades comme génétiques. Ce faisant, elle est relative à une population de contraste, ce qui a comme conséquence d'éliminer tout concept de maladie génétique en soi, mais de reconnaître un concept de maladie génétique relatif à une population donnée. Le problème consiste alors à trouver des critères non arbitraires qui permettent de définir le choix de la population de contraste. Mais en l'absence de tels critères, l'approche différentialiste donne nécessairement naissance à l'approche pragmatique qui reconnaît la dépendance des explications génétiques de la maladie aux contextes, intérêts et objectifs donnés des communautés qui cherchent et qui produisent de telles explications génétiques. L'approche pragmatique aboutit ainsi à nier la possibilité d'un concept de maladie génétique. Dès lors, on peut se poser la question d'un changement nécessaire de paradigme : plutôt que de chercher à tout prix à définir un concept de maladie génétique, ne faudrait-il pas abandonner celui-ci ?

2.3.2. Des conséquences scientifiques et sociales limitées

C'est précisément cette dernière position que nous souhaitons défendre : nous n'avons pas besoin de concept de maladie génétique. En effet, nous avons souligné dans le premier chapitre de cet exposé à quel point le concept de maladie génétique était devenu une source de confusion sur le plan scientifique, recouvrant potentiellement n'importe quelle maladie et ayant perdu toutes les caractéristiques qui ont autrefois fait sa spécificité quand le concept de maladie génétique désignait un petit groupe de maladies monogéniques mendéliennes héréditaires. Nous avons maintenant également démontré qu'aucune défense philosophique du concept de maladie génétique n'était satisfaisante : l'approche cause / condition est incompatible avec nos connaissances scientifiques actuelles, l'approche différentialiste amène inéluctablement à adopter l'approche pragmatique et celle-ci aboutit à affirmer que le concept de maladie génétique n'a pas d'intérêt.

Quelles conséquences y aurait-il à abandonner ce concept de maladie génétique ? Paradoxalement, certains pourraient s'inquiéter des conséquences pragmatiques d'un tel abandon pour les scientifiques, les médecins ou les patients. Il est vrai, comme le font justement remarquer les défenseurs de l'approche pragmatique, que les concepts médicaux jouent un rôle social dans la prise en charge thérapeutique des patients, dans la distribution des ressources publiques pour la recherche médicale, dans la perception qu'ont les malades de leur maladie.

D'ailleurs, le Projet Génome Humain s'est accompagné d'un financement spécifique dédié à l'étude de ses conséquences éthiques, législatives et sociales (*Ethical Legal and Social Implications* - ELSI). Depuis, a surgi une littérature bioéthique considérable sur des sujets aussi divers que le dépistage pré-implantatoire et le tri des embryons, la généralisation des tests génétiques, les discriminations possibles auprès des compagnies d'assurance, l'utilisation judiciaire de l'information génétique, et ainsi de suite[13]. C'est ainsi que certains auteurs ont même parlé de la nécessité de développer une « généthique » (*genethics*) : ce concept, forgé par David Suzuki et Peter Knudtson (Suzuki et Knudtson, 1989), désignerait une branche spécifique de l'éthique médicale dédiée aux problèmes posés par l'accélération des technologies de séquençage, la généralisation des tests génétiques et l'avalanche de données issues du Projet Génome Humain. Le néologisme a connu un succès suffisant pour que d'une part la revue *Cambridge Quarterly of Health Care Ethics* annonce dès 1995 l'ouverture d'une section spéciale dédiée à la généthique[14] et pour que d'autre part, la collection de référence *Companion* ait édité en 2002 un *Companion to Genethics* (Burley et Harris, 2008), ce qui comme le fait remarquer Lewens (Lewens, 2004), est tout à fait inhabituel : il est rare que la collection s'intéresse aux défis technologiques et il n'existe pas par exemple de manuel d'éthique spécifique au don d'organe. La génétique pose-t-elle des défis si spécifiques ou si globaux qu'elle impose la création d'un nouveau champ de recherche en éthique – auquel cas, il semblerait absolument nécessaire de conserver le concept de maladie génétique, en dépit des difficultés qu'il posait ?

[13] Pour une introduction à ces différents sujets, on pourra se reporter à l'ouvrage collectif *The Code of Codes – Scientific and Social Issues in the Human Genome Projet* (Kevles et Hood, 1992).

[14] Il est intéressant de noter que parmi les auteurs de cet éditorial figurent Leslie G. Biesecker et Francis S. Collins, deux personnalités de premier plan dans le Projet Génome Humain.

Plusieurs auteurs ont au contraire dénoncé cet « exceptionnalisme génétique », qui fait de l'information génétique une information tout à fait spéciale dans le champ de l'éthique (Holm, 1999 ; Richards, 2001 ; Holtzman, 2001 ; Green et Botkin, 2003 ; Ilkilic, 2009). Ainsi, Hélène Richard défend l'idée selon laquelle du point de vue des questions de dépistage prénatal, les maladies génétiques ne posent pas de problème éthique spécifique par rapport aux maladies non génétiques, déjà dépistées de façon routinière en population générale (Richard, à venir). Au-delà de cet exemple spécifique de contexte éthique où la distinction entre les maladies génétiques et les maladies non génétiques n'est pas pertinente, Peter Kakuk passe en revue les différents arguments usuels de l'exceptionnalisme génétique (pouvoir prédictif du gène, caractère unique de l'information génétique,…) et considère que ces arguments sont en grande partie dépendants d'un modèle de la maladie génétique qui est justement celui de la maladie monogénique mendélienne dans les années 1960 et autour d'un concept de gène, tout droit issu du dogme de la biologie moléculaire. Or, comme précisément ces modèles ont été remis en question, Peter Kakuk considère que les arguments sur lesquels s'est reposé l'exceptionnalisme génétique sont en grande partie dépassés et que le discours exceptionnaliste devient caricatural, ou en tout cas peu représentatif des questions éthiques que pose la généticisation des maladies en général :

> « Pour interpréter les problèmes éthiques et sociaux posés par l'information génétique, le cas des maladies monogéniques est souvent utilisé comme un modèle ou un paradigme pour juger la valeur de la prédiction génétique à un niveau plus général et pour comprendre les implications sociales plus larges des nouvelles possibilités d'acquérir une information génétique à partir d'un individu. Cette approche entraîne une mauvaise interprétation de la signification et de l'étendue des problèmes éthiques, parce que ces maladies sont très rares au niveau de la population générale et que la valeur prédictive de l'information du génotype est exceptionnelle en termes de fiabilité et de certitude. Les maladies les plus communes dans nos sociétés modernes et la plupart des traits humains sont déterminés de façon complexe, tombant ainsi en dehors du modèle génotype-phénotype des maladies monogéniques. » (Kakuk, 2008, p. 368)

Peter Kakuk ne nie pas complètement la pertinence de l'exceptionnalisme génétique, mais il considère que le discours bioéthique qui l'accompagne doit se renouveler en prenant en compte les interactions complexes qui existent dans toutes les maladies entre génome et environnement :

> « Bien qu'on parle encore beaucoup de « gène de ceci et de gène de cela », de code génétique, d'information génétique et d'empreinte génétique, ces métaphores ont perdu les fondements scientifiques qu'elles avaient autrefois. Cela laisse à la délibération bioéthique la tâche de déconstruire les problèmes éthiques qui se sont organisés autour d'une vision simple du gène qui a donné l'impression que la réflexion éthique ne voit pas les complexités de la pensée biologique. Le discours bioéthique devrait passer d'un discours scientifique génocentré à des discours plus sophistiqués et plus complexes où le développement humain est représenté comme un problèmes d'interactions complexes entre les génomes et les environnements, entre les gènes, les facteurs éducatifs, les régimes nutritionnels et d'autres ressources développementales. Si un gène est vu comme l'une des différentes ressources du développement qui forment un trait humain donné, alors à la fois la généthique et l'exceptionnalisme génétique pourraient difficilement représenter une approche justifiée pour discuter les problèmes éthiques et sociaux de la génétique. » (Kakuk, 2008, p. 373)

Il ne semble donc pas y avoir d'argument social ou éthique sérieux pour maintenir le concept de maladie génétique – au contraire, ce concept aurait tendance à entretenir une confusion sur la réalité des problèmes éthiques posés dans le contexte d'une généticisation des maladies. On pourrait d'ailleurs arguer qu'il existe d'autres concepts peut-être plus adaptés pour aborder des questions éthiques et sociales, et à vrai dire, moins ambigus à définir et plus aisés à manipuler, car dépourvus des connotations génocentristes d'un autre âge qui hantent le concept de maladie génétique : on peut penser au concept de « maladies orphelines », au concept de « maladies congénitales » ou au concept d'« anomalies chromosomiques », pour ne donner que quelques exemples. Il ne s'agit pas de substituer d'un point de vue épistémologique ces autres concepts à celui de la maladie génétique, simplement d'affirmer que le rôle social et éthique du concept de maladie génétique peut être

assumé avec d'autres termes moins ambigus et ne doit pas être utilisé pour justifier le maintien d'un concept dépassé.

2.4. Il faut abandonner le problème de la sélection causale

Nous avons montré que le concept de maladie génétique et le problème de la sélection causale étaient intrinsèquement liés. Il semble donc logique, si l'on accepte d'abandonner le concept de maladie génétique, d'abandonner également le problème de la sélection causale. En dehors de ce lien logique entre concept de maladie génétique et problème de la sélection causale, il nous semble qu'il y a quatre raisons supplémentaires de se débarrasser du problème de la sélection causale dans l'explication des maladies – raisons qui sont toutes exemplifiées par l'exposé des différentes stratégies philosophiques que nous avons proposé plus haut.

2.4.1. Résoudre le problème de la sélection causale n'est d'aucune utilité en médicine

La sélection causale est une façon relativement grossière de mesurer la responsabilité de différentes causes dans un événement donné puisque cela suppose de sélectionner une seule cause comme la plus importante dans un événement de nature multicausale. On considère souvent quand on applique ce problème à la maladie que c'est précisément ce qui nous intéresse : sélectionner le facteur causal le plus important afin de trouver le traitement le plus efficace. Ce lien entre sélection causale et efficacité thérapeutique semble assez séduisant en particulier dans un contexte de santé publique. On peut formuler au moins trois objections à cet argument.

(1) D'abord la cause principale d'une maladie n'est pas toujours la meilleure cible pour une intervention médicale. Par exemple, on considère souvent dans l'exemple de la phénylcétonurie que la mutation du gène PAH est le facteur causal le plus important. Mais quelle est l'intervention la plus efficace pour cette maladie ? Intervenir sur le gène (ce que nous ne savons pas faire pour le moment) ou prescrire un régime sans phénylalanine ? Comme nous l'avons remarqué plus haut en évoquant

la question de la manipulabilité, il semble que même si l'on disposait d'une méthode de thérapie génique permettant d'introduire une copie du gène normal, le régime sans phénylalanine resterait une méthode plus simple et plus efficace d'éviter l'apparition de la phénylcétonurie.

(2) Le second argument est qu'une maladie peut avoir différentes causes indépendantes les unes des autres. Si on prend l'exemple de l'ulcère gastroduodénal, il y a ainsi au moins deux causes principales : la majorité des cas est causée par une infection à *Helicobacter Pylori*, une bactérie susceptible de coloniser l'estomac et qu'on traite par antibiotiques, mais les autres cas sont la plupart du temps liés à une acidité excessive de l'estomac, qui sera traitée par des inhibiteurs de la pompe à protons. Ainsi, d'un patient à un autre, la cause principale d'une maladie peut varier.

(3) Le troisième argument est de dire qu'une maladie peut avoir des causes multiples et que se débarrasser de la cause dite principale (dans le cas où cette cause principale est accessible à une intervention médicale) ne signifie pas nécessairement que le malade sera guéri. Ainsi, pour reprendre l'exemple de l'ulcère, on pense aujourd'hui qu'il y a un lien entre la présence d'*Hélicobacter Pylori* et une sécrétion excessive d'acide dans certains cas d'ulcère. Cela signifie que si certains cas d'ulcères sont uniquement causés par *Hélicobacter Pylori* et que d'autres n'impliquent absolument pas *Helicobacter Pylori* mais sont seulement dus à un excès d'acidité, certains cas sont mixtes : l'infection par *Hélicobacter Pylori* entraîne une sécrétion excessive d'acidité qui elle-même majore la maladie. Il ne suffira donc pas de se débarrasser de l'infection dans ce cas, il faudra aussi traiter l'excès d'acidité.

2.4.2. Résoudre le problème de la sélection causale ne fournit pas une explication de la maladie

Sélectionner la cause la plus importante pour une maladie donnée semble impliquer une connaissance complète du contexte causal : il faudrait avoir une connaissance complète de la matrice causale pour pouvoir justifier le choix d'une population de contraste comme le suggère l'exemple de l'eau contaminée dans le village. Pourtant, résoudre le problème de la sélection causale n'implique pas nécessairement de connaître la façon dont l'ensemble des causes interagissent entre elles ou dépendent les unes des autres. En d'autres termes, résoudre le problème de la

sélection causale ne nous donne pas une explication causale de la maladie. Pire encore, sélectionner une cause parmi d'autres dans un contexte multicausal suppose que l'on traite les causes de la maladies indépendamment les unes des autres. En appelant une maladie « génétique », on implique en un sens que la façon dont les gènes et l'environnement interagissent pour causer la maladie ne mérite pas qu'on s'y intéresse. Ainsi non seulement la sélection de la cause la plus importante ne fournit pas une explication de la maladie, mais on pourrait défendre qu'elle va jusqu'à nous éloigner d'une explication causale de la maladie. On peut donner ici l'exemple de l'anémie falciforme, classiquement considérée comme une maladie génétique. Une fois qu'on a défini la maladie comme génétique en raison de l'implication de la mutation du gène HbS dans la physiopathologie de la maladie, on n'a cependant pas expliqué pourquoi certains drépanocytaires ont une symptomatologie beaucoup plus sévère que d'autres ou par quels mécanismes des modifications environnementales (climat d'altitude, infection, etc.) peuvent déclencher les différentes manifestations de la maladie. C'est pourtant cet ensemble d'informations que recherchent la plupart des patients qui apprennent qu'ils sont porteurs de l'anomalie génétique de la drépanocytose. Un phénomène similaire pourrait être discuté avec l'exemple de la mucoviscidose : si on sait que la mucoviscidose implique plus de mille mutations différentes sur le gène CFTR qui code la production des canaux chlore, on comprend encore imparfaitement le lien entre les mutations du gène CFTR et la diversité anatomo-clinique des manifestations respiratoires et pancréatiques (pour ne citer que les deux principales) de la mucoviscidose. On ne sait pas non plus comment corréler la diversité des mutations avec la diversité de la sévérité des phénotypes. Résoudre le problème de la sélection causale n'est donc ni nécessaire ni stimulant pour découvrir l'explication causale d'un phénomène.

2.4.3. Le problème de la sélection causale simplifie le contexte causal

Ce point est une conséquence de notre argument précédent. Sélectionner la cause la plus importante dans l'explication d'un phénomène implique que les interactions entre les différentes causes devraient être ignorées. De façon similaire, cela a aussi pour conséquence de simplifier la façon dont on définit l'ensemble du contexte causal. Dans notre cas, identifier une maladie comme génétique signifie

qu'on insiste sur le rôle causal des gènes au détriment du rôle de l'environnement. Les gènes sont définis de façon très floue comme n'importe quel brin d'ADN – ignorant ainsi le fait bien établi que le concept de gène soulève de sérieux problèmes de définition et que différents types de gènes ont différentes fonctions dans l'organisme, etc. Il est ainsi assez étonnant dans l'ensemble de la littérature que nous avons évoquée de constater que les anomalies chromosomiques qui impliquent la délétion ou l'ajout de plusieurs dizaines de gènes soient souvent mises sur le même plan que les mutations alléliques sur un ou plusieurs gènes. L'environnement est défini encore moins précisément – souvent comme tout ce qui n'est pas le gène ; ce qui inclut à la fois l'environnement intérieur et l'environnement extérieur, aussi bien l'épigénétique que le milieu socio-économique, aussi bien des processus cellulaires stochastiques que la toxicité d'agents mutagènes extérieurs. Formuler le problème de la maladie génétique en termes de sélection causale semble ainsi aboutir à définir une matrice causale simpliste qui se réduit à un choix binaire entre deux types de causes mal définies.

2.4.4. Le problème de la sélection causale devrait être abandonné pour interpréter de façon différente et pertinente la généticisation

Finalement, comme nous l'avons dit précédemment, une fois que la généticisation a été interprétée comme une extension du concept de maladie génétique à toute maladie et que le concept de maladie génétique a été interprété comme une instance du problème de la sélection causale, la généticisation est comprise comme l'affirmation que pour chaque maladie, les gènes sont la cause la plus importante – ce qui est une affirmation dont l'absurdité est aisée à démontrer. Mais si nous nous débarrassons du concept de maladie génétique et que nous mettons de côté le problème de la sélection causale, il devient possible de comprendre la généticisation dans un cadre interactionniste comme l'affirmation que les gènes sont toujours des facteurs causaux pertinents dans une explication de la maladie. Reste alors à savoir si nous pouvons maintenant mieux définir ce rôle causal des gènes dans toutes les maladies et lui donner un sens non trivial.

2.5. Conclusion du chapitre : changer l'approche de la généticisation

Dans ce chapitre, nous avons décrit la façon dont les philosophes analysent le problème de la définition du concept de maladie génétique d'une part et l'extension de ce concept à toute maladie, c'est-à-dire la généticisation, d'autre part. Nous avons démontré que la majorité des auteurs adoptent une formulation commune de ces difficultés en termes de sélection causale. Ce faisant, ils s'opposent à la généticisation et proposent différentes techniques ou critères pour résoudre le problème de la sélection causale. Trois stratégies sont principalement évoquées dans ce but, qui, soit échouent à résoudre le problème de la sélection causale, soit soulignent la relativité des explications génétiques de telle sorte qu'elles aboutissent à nier la possibilité de définir un concept de maladie génétique. Prenant acte de cet échec, nous proposons un changement de paradigme. Ainsi, plutôt que de combattre la généticisation afin de résoudre le problème de la sélection causale dans l'explication de la maladie et de sauver un concept peu clair de maladie génétique, dont l'utilité sociale et éthique est loin d'être évidente, nous soutenons qu'il faut abandonner le concept de maladie génétique et le problème de la sélection causale afin de donner un sens à la généticisation de la maladie et de comprendre de façon pertinente le rôle causal que jouent les gènes dans toute maladie.

Partie 2 : De la généticisation des maladies aux conditions de possibilité et aux critères d'une théorie génétique de la maladie

La première partie de ce travail nous a permis de changer de paradigme, c'est-à-dire de passer de l'examen du concept de maladie génétique à l'examen de la généticisation des maladies, de la question de savoir en quel sens une maladie est génétique à la question de savoir en quel sens toutes les maladies sont génétiques. Dans la deuxième partie de ce travail, il s'agit de poser les fondements de ce changement de paradigme, c'est-à-dire d'examiner à quelles conditions on peut interpréter la généticisation des maladies comme l'élaboration progressive au sein de la littérature biomédicale contemporaine d'une théorie génétique de la maladie. Le concept de « théorie génétique de la maladie » pose deux types de difficultés – une difficulté quant à ce qu'on entend par « génétique » (chapitre 3) et une difficulté quant à ce qu'on entend par « théorie » (chapitre 4).

Dans le chapitre 3, nous examinons la première difficulté – quel type d'interactionnisme constitue une condition nécessaire à l'élaboration d'une théorie génétique de la maladie qui ne soit ni triviale ni génocentriste ? En effet, parce que nous avons établi que la généticisation des maladies ne saurait ni se réduire à l'affirmation triviale que les gènes jouent un rôle quelconque dans toutes les maladies, ni se réduire à l'affirmation génocentriste que les gènes sont la cause principale de toutes les maladies, la généticisation des maladies comme élaboration d'une théorie génétique de la maladie ne peut faire sens qu'à condition qu'on établisse un cadre interactionniste pour l'explication de toutes les maladies.

Nous commençons par explorer l'interactionnisme de la génétique quantitative, ce que nous appelons « l'interactionnisme quantitatif », qui repose sur le concept d'héritabilité et sur une décomposition additive de la variance phénotypique entre variance génétique et variance environnementale. Cet interactionnisme quantitatif a contribué de façon décisive au cours du vingtième siècle à l'élucidation des variants génétiques impliqués dans les maladies mendéliennes comme dans les

maladies complexes. Mais ce sont précisément ces succès dans l'élucidation des variants génétiques qui ont révélé un problème majeur pour l'interactionnisme quantitatif : le problème de l'héritabilité perdue, c'est-à-dire le fait que lorsqu'on confronte l'héritabilité estimée d'une maladie donnée dans une population donnée à l'héritabilité expliquée par les variants alléliques connus pour être impliqués dans la maladie, une partie de l'héritabilité est manquante. Ce problème de l'héritabilité perdue met ainsi en valeur une limite de l'interactionnisme quantitatif, inhérent à la définition du concept d'héritabilité : l'interactionnisme quantitatif impose en effet de considérer les gènes et l'environnement comme deux causes indépendantes l'une de l'autre, dont on peut mesurer séparément l'influence causale sur un trait, et en l'occurrence, sur les maladies. De ce point de vue, l'interactionnisme de la génétique quantitative est encore tributaire du problème de la sélection causale.

Plusieurs modèles ont été proposés pour modifier le cadre interactionniste de la génétique quantitative de façon à résoudre le problème de l'héritabilité perdue. Il s'agirait de substituer à une décomposition additive de la variance phénotypique en variance génétique et en variance environnementale, une décomposition plus fine qui prenne en compte d'une part les interactions épistatiques (aussi appelées interactions « gène x gène » ou interactions GxG) et d'autre part les interactions « gène x environnement » (ou interactions GxE). Mais, les généticiens quantitatifs conçoivent le concept d'interaction « gène x environnement » dans son sens biométrique ou statistique – comme une déviation du modèle standard additif de l'héritabilité. À l'opposé de cette interprétation, les promoteurs d'une génétique du développement ont défendu une interprétation biologique du concept d'interaction « gène x environnement » et soutenu qu'il était nécessaire de progresser dans la compréhension des différents mécanismes biologiques en jeu dans ces interactions. Ce serait précisément dans l'élucidation de ces mécanismes biologiques des interactions GxE que se trouverait la clé d'une explication des événements du développement dont la maladie est un aspect parmi d'autres.

C'est pourquoi nous nous tournons vers l'examen d'une autre forme d'interactionnisme, un « interactionnisme qualitatif » ou « interactionnisme co-constructionniste », selon lequel les gènes et l'environnement co-construisent l'organisme par leurs interactions. Cet interactionnisme co-constructionniste a été

décrit par les fondateurs de la théorie des systèmes en développement autour de la thèse de la parité causale, qui réfute la prédominance causale des gènes dans la construction de l'organisme et établit que les gènes et l'environnement doivent être considérés comme des causes « sur un pied d'égalité », que l'on considère l'évolution ou le développement des organismes. Cette forme d'interactionnisme co-constructionniste est en théorie applicable à l'explication des maladies et présente l'avantage de se focaliser sur une description fine de la matrice causale d'une maladie donnée. Il reste pourtant une difficulté de taille : le risque de l'interactionnisme co-constructionniste, c'est de tenir à ce point à une description complexe des systèmes en développement que celle-ci pourrait bien perdre tout contenu informatif et nous ramener au « consensus interactionniste » et à une interprétation triviale de la généticisation des maladies. La première condition de possibilité d'une théorie génétique de la maladie est donc de s'appuyer sur un cadre interactionniste qualitatif de type co-constructionniste, qui propose une description complexe de la matrice causale d'une maladie donnée, tout en permettant de distinguer d'un point de vue heuristique le rôle des gènes du rôle des autres facteurs explicatifs d'une maladie.

Dans le chapitre 4, nous examinons la seconde difficulté – quelles sont les conditions nécessaires pour parler de *théorie* génétique de la maladie ? Notre objectif est ici de proposer une conception minimalement normative de ce à quoi pourrait ressembler une théorie génétique de la maladie : il ne s'agit pas d'aboutir à une définition unique du concept de théorie génétique, mais d'envisager le spectre des formes possibles que pourrait prendre une théorie génétique de la maladie, d'une théorie génétique *a minima* à une théorie génétique *a maxima* ou idéale.

Pour respecter cette exigence, nous proposons d'abord de nous appuyer sur une conception déflationniste du concept de théorie scientifique, définie comme une hypothèse scientifique étayée par des preuves empiriques qui permet d'unifier sous une même explication des classes de faits indépendants. Une fois cette conception déflationniste des théories scientifiques posée, nous abordons la définition du concept de théorie médicale à partir de trois questions sur la spécificité de l'objet, la forme de l'explication et le degré de précision qu'on peut attendre d'une théorie médicale. Cette interrogation nous conduit à identifier un objet spécifique des théories

médicales (la maladie) qui peut se décliner en différents objets-maladie selon que l'on s'intéresse à la maladie en général, à une classe de maladie, ou à une maladie individuelle en tant qu'elle est un prédicat nosologique (la tuberculose par exemple) ou en tant qu'elle est instanciée dans un individu donné (la tuberculose de Julie par exemple). À cette typologie d'objets correspondent différents niveaux possibles de théorie médicale, selon qu'on est dans une théorie générale de la maladie, dans une théorie d'une classe de maladie ou dans une théorie d'une maladie individuelle. Mais cette variété de niveaux d'une théorie médicale ne doit pas masquer l'existence de cibles explicatives communes (les causes, les symptômes, l'évolution et le traitement) à toute maladie.

Par ailleurs, on constate un pluralisme explicatif de fait à l'œuvre dans les sciences médicales. Ainsi, lorsqu'on cherche à expliquer une maladie, nous identifions plusieurs explications (explication épidémiologique, explication évolutionnaire et explication physiopathologique), qui sont chacune complète par rapport à elle-même mais complémentaire et dépendante des autres, et qui peuvent intervenir pour toutes les cibles explicatives de la maladie (les causes, les symptômes, l'évolution et le traitement d'une maladie). Ces différentes explications sont d'ailleurs toutes utiles à l'explication de la maladie d'un individu particulier : elles apportent des informations complémentaires qui rendent vaine toute hiérarchie entre ces formes d'explications. C'est pourquoi on peut affirmer qu'*une théorie médicale* a minima *suppose au minimum un type d'explication (évolutionnaire, épidémiologique, ou physiopathologique) d'au minimum une cible explicative (causes, symptômes, évolution, ou traitement) d'une maladie individuelle en tant que maladie-type, ou d'une classe de maladie ou de la maladie en général.* À partir de cette définition d'une théorie médicale *a minima* et en combinant ces différentes dimensions (type d'objet-maladie, niveau de la théorie, cible explicative, type d'explication), on peut donc envisager un spectre des théories médicales possibles.

Pour parvenir à la définition d'une théorie médicale *a maxima* ou théorie médicale idéale, nous identifions trois critères – la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité – qui constituent des critères de précision permettant d'évaluer et de comparer deux théories médicales en compétition. Nous appliquons ensuite ces critères d'identification et d'évaluation

d'une théorie médicale au concept de théorie génétique – comprise comme une théorie partielle de la maladie, c'est-à-dire une théorie médicale qui ne se concentre que sur un facteur explicatif de la maladie parmi d'autres, le gène. Les ambiguïtés de la définition du concept de gène, qui fait l'objet de débats considérables en philosophie de la biologie, pourraient paraître un obstacle à l'élaboration d'une théorie génétique : il nous semble au contraire que les ambiguïtés de cette définition pourraient permettre en principe à une théorie génétique d'intégrer différents types d'explication (évolutionnaire, épidémiologique et physiopathologique) et renforcent la crédibilité de notre hypothèse sur la généticisation des maladies comme théorie génétique de la maladie. De la même manière que nous avons identifié un spectre des formes possibles de théorie médicale, on peut alors identifier un spectre des théories génétiques possibles, selon que cette théorie constitue *une théorie génétique de la maladie* en général et s'appuie sur une définition du rôle des gènes dans la maladie en général ou selon qu'elle représente un assemblage hétéroclite de théories génétiques de classes de maladie (ce que nous appelons *une théorie génétique des maladies*).

À l'issue de cette deuxième partie, nous pouvons donc raffiner notre hypothèse sur la généticisation des maladies comme théorie génétique. Pour n'être ni génocentriste, ni triviale, une théorie génétique doit d'abord s'inscrire dans le cadre d'une explication interactionniste et co-constructionniste des maladies. Pour être une théorie applicable à l'ensemble des maladies, deux formes particulières de théories génétiques générales peuvent être envisagées : une *théorie génétique des maladies*, qui serait un assemblage hétéroclite de théories génétiques régionales spécifiques à chaque classe de maladie, ou une *théorie génétique la maladie*, qui s'appuierait sur une définition du rôle commun des gènes dans la maladie et serait susceptible de remettre en cause les classes de maladie, telles que nous les connaissons. La troisième partie de cette thèse consistera à tester ces deux hypothèses, à partir de deux exemples de théories génétiques, issus de la littérature biomédicale contemporaine.

Chapitre 3 : Comprendre la généticisation au sein d'une explication interactionniste de la maladie

Au terme du chapitre 2, nous avons montré que les différentes stratégies philosophiques pour faire sens du concept de maladie génétique avaient échoué de façon récurrente. Au delà des problèmes spécifiques à chaque stratégie discutée, c'est le fait même d'aborder la généticisation et la définition du concept de maladie génétique sous l'angle du problème de la sélection causale qui nous paraît en cause dans cet échec. Formuler le problème en termes de sélection causale semble en effet aller de pair avec une description simpliste de la matrice causale de la maladie, réduite à une opposition binaire entre gènes et environnement. Cette description simpliste ne permet pas de prendre en compte ni la complexité des facteurs causaux ainsi désignés (il existe différents types de gènes avec différentes fonctions, brin d'ADN et gène ne sont pas synonymes, l'environnement ne peut être défini comme tout ce qui n'est pas génétique) ni leurs interactions qui sont pourtant déterminantes dans l'apparition du phénotype final (diversité et sévérité des symptômes de la maladie) et dans la variabilité des symptômes au cours du temps et dans des environnements différents. C'est pourquoi nous nous proposons de changer radicalement de paradigme : il s'agit de trouver une interprétation sérieuse de la généticisation, c'est-à-dire de comprendre, dans un cadre ni trivial ni génocentriste, le rôle commun joué par les gènes dans toutes les maladies. Comprendre le sens du processus sociologique de la généticisation impose donc comme préalable nécessaire la définition d'un cadre interactionniste pour l'explication des maladies.

Ce changement de paradigme ne va pas sans difficultés, principalement parce que le terme d'interactionnisme est très controversé en philosophie de la biologie. En outre, il y a un sens trivial de l'interactionnisme, que tous les auteurs que nous avons abordés au chapitre précédent évoquent à un certain moment de leur argumentation sur la généticisation, et que nous appelons « interactionnisme vague », en raison de son caractère imprécis, trop généraliste et peu informatif. Cet interactionnisme consiste à affirmer que, d'une certaine façon, les gènes sont causalement impliqués dans tous les processus biologiques, y compris dans les processus pathologiques

comme les maladies, puisque l'ADN code les protéines, acteurs essentiels de tout processus biologique. Kelly Smith décrit ce problème en ces termes :

> « En outre, l'implication des gènes dans les traits biologiques est à ce point omniprésente qu'il semble difficile d'imaginer un trait dans lequel les gènes *ne soient pas* impliqués. De ce point de vue alors, tout trait serait labélisé comme génétique – une position que certains chercheurs en médecine adoptent de fait en ce qui concerne la maladie. Cependant il est difficile de voir en quoi une description de la sorte fournit une information d'une quelconque utilité, précisément parce que le trait en question est universel. Cela n'établit en rien que les gènes soient de quelque façon que ce soit *plus* importants que les autres facteurs causaux, bien que cela semble impliqué par le label « génétique ». Pour être tout à fait franc, si on se fonde uniquement sur l'implication causale, nous avons exactement les mêmes raisons de décrire tous les traits biologiques comme « fondés sur les protéines » ou « organiques ». Ces descriptions semblent futiles précisément parce qu'elles ne fournissent aucune information nouvelle. Ainsi, si les gènes doivent être sélectionnés causalement comme la meilleure explication pour certains traits, cela doit être sur la base de caractéristiques uniques de leur implication causale et non en vertu du simple *fait* qu'ils sont impliqués. » (Smith, 2007, p. 88–89)

Dans cette situation, Kelly Smith se place du point de vue de la sélection causale. Il montre que l'implication causale des gènes dans tout trait biologique ne suffit pas à sélectionner les gènes comme la cause la plus importante d'un trait. Mais son propos sert aussi le nôtre : le danger d'un interactionnisme « vague » serait de rejeter le problème de la sélection causale appliqué aux maladies génétiques seulement pour proposer une description triviale du rôle des gènes dans tous les traits biologiques qui ne nous dise rien de l'explication des maladies et du rôle spécifique des gènes dans cette explication. Pour éviter les écueils de l'interactionnisme vague, il nous faudra donc trouver les fondements de ce que nous appelons un « interactionnisme fort » qui permette de caractériser le rôle causal des gènes dans

toutes les maladies de façon plus spécifique. C'est dans ce contexte que nous allons aborder deux conceptions de l'interactionnisme : un interactionnisme que nous appelons « quantitatif » et un interactionnisme qui a successivement été qualifié de « constructionniste » ou de « co-constructionniste ».

L'interactionnisme que nous appelons « quantitatif » parce qu'il a été développé essentiellement par des biologistes issus de la génétique quantitative, s'appuie entre autres sur le concept d'héritabilité et consiste à partitionner de façon additive la variance phénotypique en variance génétique (elle-même additive) et en variance environnementale. Cet interactionnisme a joué un rôle majeur dans « la chasse aux gènes » (Jordan, 2006) et dans l'élaboration des modèles proposés pour décrire l'architecture allélique des maladies. Sans remettre en cause l'extraordinaire fécondité de ce modèle dans l'identification des variants alléliques impliqués dans les maladies, nous démontrons que ce modèle reste intrinsèquement tributaire du problème de la sélection causale et de la recherche indépendante d'effets génétiques et d'effets environnementaux. Ce point a peu posé problème aux généticiens quantitatifs mais se fait de plus en plus pressant, à présent que ces mêmes méthodes de la génétique quantitative ont mis en évidence le problème de l'héritabilité perdue (pour de nombreux traits, une fraction non négligeable de l'héritabilité estimée n'est pas expliquée par la variance additive des allèles impliqués dans le trait en question) et que le modèle de l'héritabilité est remis en cause par la prise en compte des interactions épistatiques (ou interactions gène x gène ou interactions GxG) et des interactions épigénétiques (ou interactions gène x environnement ou interactions GxE).

Ces critiques nous amènent naturellement à nous tourner vers la question de l'interactionnisme « constructionniste » ou « co-constructionniste », tel qu'il est défendu par les partisans de la théorie des systèmes en développement. L'interactionnisme co-constructionniste repose entre autres sur la thèse de la parité causale selon laquelle il faut traiter les gènes et l'environnement comme des causes sur un pied d'égalité dans la construction du développement et de l'évolution des organismes. Cependant, on a souvent reproché à cette forme d'interactionnisme sa dimension programmatique et peu applicable. Nous examinerons la légitimité de cette critique de l'interactionnisme co-constructionniste dans le cadre de l'explication des

maladies. En utilisant ces deux conceptions de l'interactionnisme, l'objectif de ce chapitre est donc de préciser quel cadre interactionniste peut servir de préalable à une interprétation sérieuse de la généticisation. Au terme de ce chapitre, nous formulerons l'hypothèse que la généticisation, loin de se réduire à sa description sociologique initiale ou à la reconnaissance triviale du rôle des gènes dans l'apparition de tous les traits biologiques, peut peut-être être comprise comme le reflet de l'élaboration d'une théorie génétique de la maladie.

3.1. Du concept d'héritabilité au cadre interactionniste de la génétique quantitative

Nous avons évoqué l'écueil d'un interactionnisme vague. Il semble dès lors logique, pour qualifier notre explication interactionniste des maladies, de nous tourner vers la génétique quantitative qui, à travers des concepts comme celui d'héritabilité – concepts qui sont applicables à tout trait et à toute maladie – se propose de déterminer la responsabilité causale des gènes et de l'environnement dans la genèse d'un trait biologique ou d'une maladie. Certains généticiens quantitatifs, à l'instar de Neven Sesardic, qui considèrent le terme d'interactionnisme comme une insulte, pourraient sans doute s'offusquer qu'on qualifie d'interactionniste le cadre de pensée défini par les généticiens quantitatifs. Ce terme ne nous paraît pas impropre dès lors qu'on accepte pour prémisse qu'il y a différents types d'interactionnisme et que peut être considérée comme « interactionniste » au sens large toute tentative de déterminer le type de relations qu'entretiennent les gènes et l'environnement dans la genèse d'un trait ou d'une maladie. Par ailleurs, la tradition de la génétique quantitative a joué un rôle majeur dans la conceptualisation des maladies complexes, dans l'identification des gènes associés aux maladies et dans la formulation d'hypothèses sur l'architecture allélique des maladies. Il eut été absurde d'aborder le rôle commun joué par les gènes dans les maladies sans aborder, même superficiellement, cette tradition de recherche. Nous commencerons donc par revenir sur les origines et la définition du concept d'héritabilité, concept central de la génétique quantitative. Puis nous aborderons la question des méthodes d'identification des variants alléliques associés aux maladies. Enfin, nous verrons

comment la rencontre entre le concept de l'héritabilité et les résultats des associations pangénomiques a créé le problème de « l'héritabilité manquante » et comment ce problème de l'héritabilité manquante a attiré l'attention sur des phénomènes connus depuis longtemps mais souvent laissés de côté par la génétique quantitative et que sont la prise en compte des interactions gène x gène et des interactions gène x environnement. Nous conclurons sur les raisons pour lesquelles, en dépit de sa fécondité, le cadre interactionniste mis en place par les généticiens quantitatifs ne peut convenir à la formulation d'une explication du rôle commun des gènes dans la survenue et l'évolution des maladies.

3.1.1. Origine et définition du concept d'héritabilité

Si le terme « héritabilité » a été utilisé la première fois en 1940 (Lush, 1940), il faut remonter à la controverse entre mendéliens et biométriciens et à sa résolution partielle par Fisher (Fisher, 1918) pour bien comprendre la signification de ce paramètre. La controverse entre biométriciens et mendéliens est bien connue (Gayon, 1992, 2000) : alors que les mendéliens s'appuient sur les lois de Mendel pour comprendre l'évolution discrète de la variation par saut (donnant naissance au mutationnisme), les biométriciens réfutent cette conception, au motif qu'elle est incompatible avec leurs observations biométriques d'une variation continue des caractères phénotypiques (donnant lieu au gradualisme). Dans son fameux article de 1918, Fisher s'appuie effectivement sur ce point : si les lois de Mendel expliquent l'évolution des caractères discrets ou qualitatifs (des caractères tels que le phénotype mutant/le phénotype sauvage, avoir/ne pas avoir les yeux bleus), elles ne peuvent expliquer les observations biométriques accumulées sur les variations des caractères continus (tels que la taille humaine par exemple) qui semblent suivre une loi normale, c'est-à-dire une courbe de Gauss et sur lesquels l'influence de l'environnement est patente. L'objectif de Fisher est donc de trouver une méthode statistique pour comprendre cette distribution statistique des caractères à variation continue et pour déterminer la part des gènes et la part de l'environnement dans la distribution de cette variation. C'est ainsi qu'il introduit le modèle infinitésimal qui suppose qu'un caractère quantitatif est déterminé par un nombre infini de caractères mendéliens,

chacun de ces facteurs mendéliens ayant un effet additif minime sur le phénotype global.

> « L'hypothèse la plus simple, et celle que nous allons examiner, est que des caractères tels que la taille sont déterminés par un large nombre de facteurs mendéliens et que l'importante covariance entre les enfants issus de mêmes parents est due à la ségrégation de ces facteurs pour lesquels les parents sont hétérozygotes. À partir de cette hypothèse, nous essaierons de déterminer à quel point la variance, dans ses différentes composantes mesurables, au-delà de ce qui est indiqué par la corrélation fraternelle, est due aux facteurs innés et hérités. » (Fisher, 1918, p. 135).

À partir de ce modèle, Fisher introduit le concept de variance phénotypique. La variance est une mesure statistique servant à caractériser la dispersion des valeurs d'une population autour de la moyenne des valeurs de cette population. La variance phénotypique est donc accessible à l'observation : il suffit de faire le relevé quantitatif d'un caractère dans une population – par exemple de mesurer la taille de chacun des individus d'une population donnée. Pour déterminer la part de cette variance phénotypique due aux facteurs mendéliens hérités de celle due aux facteurs environnementaux, Fisher fait l'hypothèse qu'il est possible de décomposer additivement la variance phénotypique (V_P) entre la variance génotypique (V_G) et la variance environnementale (V_E), ce qui s'écrit sous la forme :

$V_P = V_G + V_E$

A ce stade, Fisher introduit une seconde supposition, selon laquelle on peut décomposer additivement la variance génotypique entre trois types de variances : la variance additive (V_A) qui représente les effets additionnés de chaque gène impliqué dans la détermination du caractère, la variance due aux interactions entre les allèles comme l'effet de dominance (V_D), la variance due aux interactions entre les gènes, autrement appelées interactions épistatiques (V_I). On peut donc maintenant écrire :

$V_P = V_G + V_E$

$V_G = V_A + V_D + V_I$

L'héritabilité est alors définie comme la part de variance phénotypique due à la variance génotypique héritée par opposition à la variance environnementale. On peut donc écrire :

$H^2 = V_G / V_P$

Chez les êtres humains, lors de la reproduction sexuée, la contribution de chaque parent se limite à la transmission d'un gamète haploïde (contenant la moitié des chromosomes parentaux), issu de la méiose et résultant d'une recombinaison. Les interactions génétiques sont donc rompues et chaque allèle est transmis indépendamment. Ainsi, dans un gamète haploïde, la variation due à l'effet de dominance et la variation due à l'effet des interactions épistatiques ne sont pas transmises à la descendance – seule la variance additive est transmise. On peut donc à présent distinguer deux concepts : l'héritabilité au sens large (notée H^2) et l'héritabilité au sens strict (notée h^2). L'héritabilité au sens large prend en compte la variance additive, la variance due à l'effet de dominance et la variance due aux interactions épistatiques ; elle mesure donc la part de variance phénotypique dans une population dues aux différences génétiques, indépendamment de l'origine de cette différence génétique. L'héritabilité au sens strict ne prend en compte que la variance additive, c'est-à-dire uniquement l'effet additif de plusieurs gènes impliqués dans la détermination des caractères. On a donc :

$H^2 = (V_A + V_D + V_I) / V_P$

$h^2 = V_A / V_P$

Régulièrement employée dans la littérature scientifique, l'héritabilité a été fréquemment l'objet de critiques ou de malentendus – peut-être justement à cause de la variété des traits concernés par ce type d'analyse et de l'utilisation de cette méthode en génétique du comportement. Ainsi, la publication d'un article de Jensen affirmant la haute héritabilité du Q.I. (Jensen, 1969) a provoqué une controverse importante, portant à la fois sur la fiabilité des tests de Q.I. pour évaluer l'intelligence, mais aussi sur la définition de l'héritabilité, sur les méthodes qui servent à l'estimer, et sur les implications sociales d'une telle affirmation en termes de politiques publiques éducatives (Gould, 1981 ; Lewontin, 1974). C'est pourquoi, avant de préciser les rapports entre héritabilité et architecture allélique de la maladie, il nous faut clarifier trois malentendus au sujet du concept d'héritabilité : (1) l'héritabilité n'est jamais

l'héritabilité d'un trait, c'est un paramètre statistique relatif à une population donnée, (2) une héritabilité élevée ne nous dit rien du caractère génétiquement déterminé du trait, (3) la définition de l'héritabilité ne doit pas être confondue avec les méthodes d'estimation de l'héritabilité.

3.1.2. L'interprétation du concept d'héritabilité

3.1.2.1. L'héritabilité est un paramètre statistique relatif à une population et à un environnement donnés

Il est essentiel de bien comprendre que l'héritabilité est un paramètre statistique dépendant d'une population donnée. Il n'est pas possible d'extrapoler les résultats de l'héritabilité d'un caractère dans une population donnée pour définir l'héritabilité d'un caractère en général.

En effet, si on reprend la définition de l'héritabilité (au sens strict ou au sens large), on constate que chaque paramètre entrant dans la définition de l'héritabilité est dépendant de la population choisie. Par exemple, la variance génétique dépend à la fois de la ségrégation dans une population donnée des allèles qui influencent le trait, de la fréquence des allèles, de la taille des effets des variants sur le phénotype final et du mode d'action des gènes. Par ailleurs, la variance environnementale peut varier en fonction des populations : si on prend l'exemple de l'héritabilité de la taille des épis de blé, il est évident qu'on n'obtiendra pas les mêmes résultats en fonction du lieu et du climat local de culture. L'héritabilité peut également différer entre les sexes ou selon les époques de la vie. Chaque estimation de l'héritabilité doit donc théoriquement être comprise comme strictement contingente à une population donnée. Théoriquement, car « en pratique », certains traits comme l'héritabilité de la taille humaine par exemple, ne varient que très peu d'une population humaine à une autre, voire même d'une espèce animale à l'autre.

L'autre point fondamental à retenir du caractère populationnel du concept d'héritabilité est que l'héritabilité d'un caractère ne nous dit rien de l'influence du génotype d'un individu particulier sur son phénotype. L'héritabilité d'un trait dit simplement quelle proportion des différences entre des individus d'une population donnée peut être attribuée à des différences génétiques.

3.1.2.2. Une héritabilité élevée n'implique pas la détermination génétique du trait

Ce point est fondamental – c'est le sens d'une partie de la critique du concept d'héritabilité telle qu'elle a été proposée par Lewontin (Lewontin, 1974). Critiquant les analyses de Jensen sur l'héritabilité du Q.I., Lewontin conteste tout particulièrement le fait qu'une haute héritabilité du Q.I. devrait conduire à renoncer aux politiques publiques éducatives, en partant du principe qu'il n'y a qu'une faible possibilité de changer le trait par des modifications environnementales. En effet, comme l'affirme Lewontin, une estimation élevée de l'héritabilité d'un trait peut avoir deux explications : (1) cela peut effectivement signifier que le caractère est génétiquement déterminé ou faiblement sensible aux changements environnementaux (2) mais cela peut aussi signifier que les génotypes de la population en question sont très différents pour un même environnement. Inversement, si on prend un trait avec une héritabilité faible, cela peut signifier que le caractère est principalement déterminé par l'environnement ou bien que dans un même environnement, la population choisie est homozygote pour ce gène. L'estimation de l'héritabilité ne peut choisir entre ces deux explications. On peut donc affirmer comme Lewontin que le modèle linéaire est une analyse locale : *« Il donne un résultat qui dépend de la distribution particulière des génotypes et des environnements dans un échantillon de population particulier. C'est pourquoi le résultat de l'analyse est historiquement limité (c'est-à-dire du point de vue spatiotemporel) et n'est pas une affirmation générale sur des relations fonctionnelles. »* (Lewontin, 1974, p. 404)

3.1.2.3. La définition de l'héritabilité ne doit pas être confondue avec les méthodes d'estimation de l'héritabilité

L'héritabilité est définie comme un rapport de variances, mais l'estimation de l'héritabilité peut emprunter différentes méthodes (Tenesa et Haley, 2013). La méthode la plus employée pour estimer l'héritabilité a été proposée par Fischer dans ce même article de 1918 et s'appuie sur la ressemblance entre apparentés, c'est-à-dire sur la corrélation des traits portés par les parents et par leurs enfants et est estimée

par un modèle de régression linéaire[15]. Il est important de comprendre que dans un tel modèle, seule la variance phénotypique est directement observable et mesurable : la variance génétique est ensuite estimée à partir d'une comparaison entre la ressemblance mesurée expérimentalement pour le caractère considéré et la ressemblance théorique entre apparentés à laquelle on peut s'attendre en raisonnant sur le nombre d'allèles partagés par les parents et les enfants. La différence entre ressemblance théorique et ressemblance mesurée permet de mesurer la part de variance phénotypique due à la variance environnementale et par soustraction, de déduire la part de variance phénotypique due à la variance génétique. Pour obtenir des résultats statistiquement pertinents, il faut donc répéter ce type de mesure sur un échantillon important : la fiabilité de l'estimation de l'héritabilité dépend en partie de l'importance de l'échantillon choisi. Pour obtenir ce type d'estimation, différentes méthodes expérimentales peuvent être utilisées : des comparaisons intrafamiliales entre parents et enfants, des comparaisons entre jumeaux monozygotes (partageant donc le même patrimoine génétique) et jumeaux dizygotes (partageant seulement la moitié de leur patrimoine génétique), etc. Chaque méthode a des risques de biais : par exemple, si la ressemblance entre parents et enfants est partiellement due à l'effet d'un environnement partagé, l'estimation de l'héritabilité sera faussée à la hausse (l'estimation de l'héritabilité du caractère sera plus élevée que l'héritabilité réelle du caractère). Notons qu'en parallèle de cette méthode traditionnelle de la ressemblance entre apparentés, d'autres méthodes ont été développées avec l'explosion des techniques de génétique moléculaire et reposent par exemple sur l'analyse des QTL (*Quantitative Trait Locus*), qui sont des gènes polymorphes dont la variation allélique influe sur la variation phénotypique d'un caractère quantitatif puis sur l'analyse des SNP (*Single Nucleotide Polymorphism*). Ces nouvelles méthodes d'estimation de l'héritabilité – qui ne remettent cependant pas en cause la définition de l'héritabilité donnée plus haut – permettent par exemple d'évaluer l'héritabilité d'un trait dans une population dont on ne connaît pas les relations de parenté (Visscher *et al.,* 2008 ; Zaitlen et Kraft, 2012) .

[15] Pour une présentation synthétique de la méthode de la ressemblance entre apparentés, on pourra se reporter à l'annexe 2 du livre « Darwinian populations and natural selection » de Peter Godfrey-Smith (Godfrey-Smith, 2009, p. 168 - 172).

3.2. Le problème de l'héritabilité perdue dans l'explication des maladies

L'héritabilité, dont les fondements conceptuels ont été introduits par Fischer au début du siècle, permet de mesurer la part de variance génétique dans la variance phénotypique d'un trait dans une population donnée. De façon indépendante, plusieurs méthodes ont été successivement mises au point pour identifier quels sont les variants alléliques responsables de la variance génétique dans la variation phénotypique d'une maladie donnée. Or lorsque l'on confronte l'estimation de l'héritabilité d'une maladie d'une part et la variance additive des variants alléliques impliqués dans la maladie en question, il y a une discordance significative. Autrement dit, une fraction de l'héritabilité reste inexpliquée par la variance additive des variants alléliques impliqués dans le trait ou la maladie en question : c'est le problème de l'héritabilité perdue. Pour bien comprendre les enjeux de ce problème de l'héritabilité perdue et la façon dont il questionne le concept même d'héritabilité – pierre angulaire de l'interactionnisme quantitatif, il est important de comprendre quelles sont ces méthodes d'identification des variants alléliques, et en particulier la méthode des associations pangénomiques.

3.2.1. Méthodes d'identification des gènes de la maladie – des succès des maladies monogéniques au « cauchemar » des maladies complexes

Nous ne pouvons faire ici une histoire détaillée des méthodes d'identification des variants alléliques impliqués dans les maladies : cette histoire complexe d'un point de vue technique a été faite ailleurs (Jordan, 2006) et nous en rappelons les principales étapes en annexe (Annexe 1 – Méthodes d'identification des gènes de la maladie – des succès des maladies monogéniques aux cauchemars des maladies complexes). Pour notre propos, il suffit cependant de bien comprendre que les méthodes dites de la « génétique inverse » (Jordan, 1988, 2006) qui ont été mises en place avec succès dans les années 1980 et 1990 pour identifier les variants impliqués dans les maladies mendéliennes monogéniques, se sont révélées très difficiles à transposer pour les

maladies multifactorielles polygéniques complexes (Botstein et Risch, 2003) – au point que certains auteurs parlent de l'élucidation de la génétique des maladies complexes comme un « cauchemar du généticien » (Feingold, 2005).

Plusieurs raisons détaillées en annexe expliquent cette difficulté. Une première méthode employée dans l'analyse des maladies mendéliennes, l'analyse de liaison génétique, est dite « paramétrique » : elle suppose de connaître un certain nombre d'informations quant au mode de transmission de la maladie. Or, précisément, le mode de transmission des maladies multifactorielles est difficile à déterminer avec précision et pour y parvenir, il faut à la fois observer un grand nombre de familles (où il n'est pas toujours évident de distinguer les cas de maladie dus à des mutations sporadiques de ceux dus à des mutations héréditaires) et prendre en compte les phénomènes de pénétrance incomplète (on risque d'exclure de l'étude de ces grandes familles des patients sains qui sont peut-être pourtant porteurs d'un allèle de susceptibilité à la maladie qui ne s'exprimerait pas), des phénomènes comme l'hétérogénéité génétique (des gènes différents peuvent donner la même maladie) ou l'existence de gènes modificateurs. Pour pallier ces inconvénients, une variante de l'analyse de liaison paramétrique utilisée pour les maladies mendéliennes a été développée pour les maladies polygéniques : les analyses de liaison non paramétriques, telles que la méthode des paires de germains atteints. Ces méthodes ont cependant leurs limites (Boehnke, 1994). Outre qu'elles reposent sur une estimation simplifiée des paramètres impliqués dans les études de liaison, elles nécessitent également des échantillons très larges, particulièrement lorsqu'il s'agit de rechercher des variants alléliques communs ayant un effet mineur sur le phénotype global de la maladie (Hirschhorn et Daly, 2005).

Étant donné les difficultés que pose l'analyse de liaison génétique dans l'étude des maladies polygéniques, d'autres méthodes d'identification des variants alléliques utilisées pour les maladies mendéliennes comme l'approche gène-candidat, ont été employées. La première variante de cette approche utilise des informations sur la physiopathologie de la maladie ou sur des gènes similaires identifiés dans des pathologies comparables chez d'autres espèces pour comparer des séquences entières d'une même région ou d'un même gène entre des individus atteints par la maladie et des individus sains. Cependant, même avec la diminution des coûts et l'augmentation

de la rapidité des techniques de séquençage, ce type d'étude reste coûteux – ce qui explique que les chercheurs impliqués se limitent généralement à séquencer un ou deux gènes d'intérêt, au risque de ne pas parvenir à tirer des conclusions significatives.

Une autre méthode pour adapter les études de gènes candidats aux maladies complexes et qui est moins coûteuse que le reséquençage de gènes candidats est l'étude d'association de gènes candidats. Les études d'association en général utilisent le principe du déséquilibre de liaison. Qu'est-ce que le déséquilibre de liaison ? Le déséquilibre de liaison mesure la fréquence avec laquelle les allèles de deux polymorphismes se retrouvent sur le même chromosome. Il est affecté par la recombinaison méiotique et diminue proportionnellement à la distance séparant les polymorphismes. Quand il existe un déséquilibre de liaison entre deux allèles, cela signifie qu'il faudra encore un grand nombre de recombinaisons méiotiques (c'est-à-dire un grand nombre de générations) pour que les deux polymorphismes en question soient en équilibre de liaison, c'est-à-dire pour qu'il y ait la même probabilité pour que les deux allèles coségrégent que pour qu'ils ségrégent séparément durant la méiose. Les études d'association différent significativement des études de liaison génétique : alors que les études de liaison portent sur l'étude de liaisons génétiques au sein de grandes familles de patients atteints par une maladie donnée, les études d'association génétique portent sur des comparaisons entre une cohorte de patients atteints et une cohorte de patients sains, sans qu'il y ait nécessairement des relations de parenté entre les parents atteints. Les études d'association ont une meilleure puissance et une plus grande précision lorsque les polymorphismes que l'on cherche à détecter ont un risque relatif faible (c'est-à-dire un effet modeste sur le phénotype) ce qui est le cas dans les maladies complexes. Dans le cadre de l'approche gène-candidat, les études d'association consistent à comparer la distribution des allèles ou des génotypes d'un marqueur entre une collection de patients atteints de la maladie et une cohorte de témoins. L'augmentation statistiquement significative de la fréquence d'un allèle du marqueur indique que le variant testé (méthode d'association directe) ou qu'un variant en déséquilibre de liaison avec le marqueur testé (méthode d'association indirecte) est un allèle de susceptibilité de la maladie. Comme les études d'association de gènes candidats utilisent quelques variants communs, au lieu de reséquencer des gènes entiers, elle est moins coûteuse que la méthode de reséquençage des gènes

candidats. Néanmoins ces études d'associations de gènes-candidats nécessitent d'avoir prédit correctement l'identité du ou des gènes candidats, en général en se fondant sur des hypothèses physiopathologiques ou bien en utilisant des gènes candidats déterminés dans une précédente étude. Ce type d'études d'associations limités à des gènes-candidats pose donc au moins trois types de problème : (1) elles ne peuvent être mises en place que si l'on dispose de connaissances physiopathologiques préalables sur la maladie (2) même si elles aboutissent à identifier quelques mutations délétères sur les gènes-candidats, elles ne garantissent pas l'identification de tous les facteurs génétiques impliqués dans la maladie – seulement d'une fraction d'entre eux (3) enfin, ces études sont relativement difficiles à répliquer à cause de problèmes d'échantillonnages assez complexes sur lesquels nous ne nous attarderons pas ici mais qui complexifient encore l'analyse des résultats de ces études d'association génétique (Cordell et Clayton, 2005 ; Tabor, Risch et Myers, 2002).

3.2.2. Les associations pangénomiques, l'hypothèse « variant commun-maladie commune »

Ce sont ces difficultés à élucider la génétique des maladies complexes, alors même que celles-ci sont considérées comme un problème majeur de santé publique, qui explique l'enthousiasme déclenché par la mise en place des études d'association pangénomiques. Bien que les études d'association pangénomiques aient été envisagées très tôt (Risch et Merikangas, 1996), leur mise en place effective n'a pu se faire que dans les années 2005 pour différentes raisons techniques. En effet, si les associations pangénomiques sont dites ne reposer sur aucune hypothèse biologique (on parle aussi de « méthode agnostique ») – par opposition aux méthodes d'analyse de liaison qui présupposent le mode de transmission de la maladie et par opposition aux méthodes « gène-candidat » qui présupposent une connaissance des bases physiopathologiques de la maladie, elles sont en faites largement dépendantes d'une hypothèse développée à la fin des années 1990 (Lander, 1996 ; Reich et Lander, 2001) et qui a dominé la génétique des maladies complexes, appelée « l'hypothèse variant commun, maladie commune ». Selon cette hypothèse, les maladies communes ne sont pas dues à une infinité de variants alléliques à faible effet sur le phénotype (ce qui correspond au modèle infinitésimal de Fischer) mais à un nombre modeste de variants

communs (dont la fréquence dans la population serait supérieure à 1%). Aucun de ces variants ne suffit à provoquer la maladie (ils sont rarement délétères) ce qui expliqueraient leur maintien dans la population générale (s'ils étaient délétères, on pourrait s'attendre à ce que la sélection naturelle aboutisse à leur élimination). Paradoxalement, même les principaux défenseurs de cette hypothèse ont toujours reconnu que cette hypothèse manquait de preuve empirique (Reich et Lander, 2001) – mais c'est l'hypothèse la plus logique pour expliquer à la fois le caractère commun des maladies, la complexité de la détermination génétique de ces maladies et le maintien de ces variants dans la population générale. Si les variants responsables des maladies communes étaient au contraire rares mais avec un effet très important sur le phénotype global, les maladies communes seraient moins fréquentes et la présence d'un allèle muté dans une famille donnerait un modèle de transmission très proche de l'hérédité mendélienne. Ajoutons à ces raisons épidémiologiques que l'hypothèse « variant commun-maladie commune » a par ailleurs bénéficié de quelques exemples bien établis, comme le variant ApoE dans la maladie d'Alzheimer qui ont contribué à en faire un modèle admis par tous.

L'hypothèse « variant commun, maladie commune » a encouragé l'idée qu'il serait possible d'élargir les études d'association au-delà d'un petit groupe de gènes candidats et au génome tout entier, à condition de disposer d'une base de données de marqueurs de déséquilibre de liaison suffisante pour couvrir la totalité du génome. On a découvert par ailleurs au début des années 2000 que le déséquilibre de liaison ne diminuait pas de façon graduelle au cours des générations comme on le pensait mais par sauts brusques. Le génome est donc organisé en blocs, appelés « haplotypes », qui mesurent en moyenne de 10 à 20 kilobases et qui sont séparés par de petites régions d'environ 1 à 2 kilobases qui ont un taux de recombinaison élevé (raison pour laquelle on parle de « points chauds de recombinaison ») et dont la structrure haplotypique sous-jacente est très similaire au sein de toutes les populations humaines. À partir de ces observations, on a calculé qu'il « suffirait » de disposer de 500 000 polymorphismes nucléotidiques simples ayant une fréquence commune (supérieure à 5%), à raison d'un SNP marqueur toutes les cinq mille bases pour avoir une couverture suffisante du génome et développer des études d'associations sur le génome entier. Cet objectif des 500 000 polymorphismes nucléotidiques simples communs est devenu

celui du projet HapMap, terminé en 2004 et publié en 2005 (The International HapMap Consortium, 2005). En parallèle de ces avancées, les techniques de génotypage se sont améliorées (augmentation de la rapidité et baisse du coût) : c'est ainsi que deux laboratoires proposent en 2004 une nouvelle génération de puces à ADN permettant d'analyser 500 000 SNP simultanément en une seule expérience et pour un coût d'une centaine d'euros. Tous les ingrédients sont alors réunis pour permettre des études d'associations sur le génome entier, aussi appelées associations pangénomiques.

Les associations pangénomiques se décomposent en plusieurs étapes : (1) il s'agit d'abord de constituer deux groupes – un groupe d'individus atteints par la maladie (le groupe des cas) et un groupe d'individus sains (groupe témoin) dont les caractéristiques générales sont aussi proches que possible de celles du groupe d'individus atteints (distribution de l'âge, du sexe, mode de vie aussi proches que possible dans les deux groupes). Précisons que pour des raisons de puissance statistique, il est nécessaire d'avoir des groupes très importants. (2) Une fois ces deux groupes sélectionnés, il faut analyser à l'aide des puces à ADN les 500 000 SNP de chacune de ces personnes, c'est-à-dire déterminer pour chacun de ces polymorphismes quels allèles sont présents sur les deux chromosomes de chaque individu. (3) On procède alors à une comparaison des résultats entre les deux groupes afin de déterminer si certains SNP sont sur-représentés ou sous-représentés dans l'un des deux groupes. A ce stade, il faut préciser un point important : il n'est pas sûr que les SNP sur-représentés ou sous-représentés dans le groupe des patients atteints de la maladie soient des variants de susceptibilité de la maladie. C'est parfois le cas, mais plus souvent ces SNP désignent seulement une région d'exploration dans laquelle se trouvent les « vrais » variants de susceptibilité ou des séquences régulatrices influençant la maladie qui sont en déséquilibre de liaison avec les SNPs sous ou sur-représentés et qu'il va falloir ensuite identifier par une étude approfondie de la région en question (4). L'immense différence avec les études de liaison ou les études d'association « gène-candidat », c'est qu'il n'est pas nécessaire de connaître ni le mode de transmission de la maladie, ni aucune information sur la physiopathologie de la maladie pour rechercher les variants – il n'y a donc pas de biais dépendants d'assomptions spécifiques sur la maladie d'intérêt.

Les premières études d'association pangénomiques ont eu lieu en 2005 et en 2006 (Dewan *et al.*, 2006 ; Klein *et al.,* 2005), mais la plupart des auteurs qui s'intéressent à l'histoire des associations pangénomiques (Visscher *et al.,* 2012) retiennent la date de 2007, lorsque le *Wellcome Trust Case Control Center* a publié les résultats d'une étude d'association pangénomique portant sur plusieurs maladies communes (comprenant entre autres la maladie de Crohn, la polyarthrite rhumatoïde, la maladie coronarienne, le diabète de type 1 et le diabète de type 2), quatorze mille cas et trois mille contrôles (Burton *et al.,* 2007*)*, en raison de la rigueur méthodologique employée. Entre les années 2007 et 2012, les associations pangénomiques se sont multipliées de façon exponentielle, comme le montre la Figure 7, issue du catalogue des associations pangénomiques publiées que l'on trouve sur le site de l'Institut National de Recherche sur le Génome Humain.

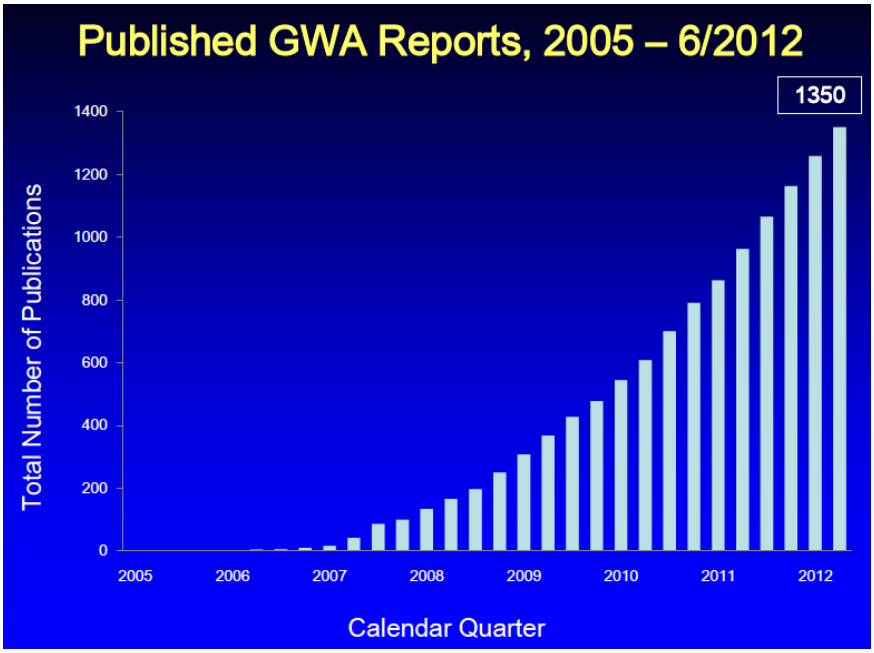


Figure 7 : Nombre d'études d'association pangénomiques publiées entre 2005 et juin 2012 - Sources : site du catalogue des études d'association pangénomiques publiées - rapport de Teri Manolio : http://www.genome.gov/gwastudies/ (consulté le 6 mars 2013)

Les associations pangénomiques ont certainement permis à la génétique des maladies complexes de progresser. C'est ainsi que Peter Visscher résume les cinq dernières années, en évoquant la découverte de 2000 nouveaux loci associés de façon significative et robuste à un ou plusieurs traits et correspondant donc à « 2000

nouvelles pistes biologiques » (Visscher *et al.,* 2012, p. 9–10). Cependant, en dépit des nombreux espoirs soulevés et de ces succès, les associations pangénomiques ont également entraîné de nombreuses déceptions (Klein, 2012 ; Ku *et al*., 2010 ; Visscher et *al*., 2012 ; Zaitlen et Kraft, 2012). Nous ne discuterons pas des débats méthodologiques sur le type de SNP utilisés ou sur les problèmes de puissance statistique.[16] Mais nous voulons aborder un problème qui nous ramène là où nous avons commencé ce chapitre – celui de « l'héritabilité perdue » ou « héritabilité manquante ». Ce phénomène désigne le fait que lorsqu'on compare l'influence des variants alléliques (identifiés par l'utilisation de marqueurs dans des associations pangénomiques) sur un caractère ou une maladie donnée aux estimations d'héritabilité issues des comparaisons de ressemblance entre apparentés ou des nouvelles méthodes moléculaires, ces variants n'expliquent qu'une petite part de l'héritabilité du trait en question – laissant de côté une part de l'héritabilité du trait dite « perdue » ou « manquante ». Plus formellement, les généticiens définissent la proportion de l'héritabilité au sens strict d'un trait expliquée par un ensemble de variants génétiques connus ($\pi_{\text{expliquée}}$) de la façon suivante :

$$\pi_{\text{expliquée}} = h^2_{\text{connue}} / h^2_{\text{totale estimée}}$$

où le numérateur h^2_{connue} correspond à la proportion de la variance phénotypique expliquée par les effets additifs des variants connus (et qui est directement observable expérimentalement) et où le dénominateur $h^2_{\text{totale estimée}}$ est la proportion de la variance phénotypique attribuable aux effets génétiques additifs et qui est estimée selon les méthodes que nous avons précédemment passées en revue. L'héritabilité manquante correspond donc au pourcentage de l'héritabilité totale telle qu'elle a été estimée qui n'est pas expliqué par les variants génétiques qui ont été identifiés, que ce soit par des approches gènes-candidats, ou par des associations pangénomiques :

> « La justification derrière les associations pangénomiques est l'hypothèse « maladie commune, variant commun » qui suppose que les maladies communes sont attribuables en partie aux variants alléliques présents dans plus de 1 à 5% de la population. Elles ont été possibles grâce au

[16] (Pour une analyse accessible de ces problèmes, voir McCarthy *et al.*, 2008)

développement commercial des « puces à SNP » ou matrices, qui capturent la plus grande partie, et non pas la totalité de la variation commune dans le génome. Bien que l'architecture allélique de certaines maladies, en particulier la dégénérescence maculaire liée à l'âge, reflète majoritairement les contributions de plusieurs variants de large effet (définis de façon vague ici comme ceux qui augmentent le risque de contracter la maladie par deux ou plus), la plupart des variants communs, qu'ils soient pris individuellement ou de façon combinée, n'expliquent qu'une petite part de l'augmentation du risque (entre 1,1 et 1,5 fois le risque) et expliquent seulement une petite proportion de l'héritabilité – la portion de la variance phénotypique dans une population qui est attribuable à des facteurs génétiques additifs. Par exemple, au moins 40 loci ont été associés avec la taille humaine, un trait complexe classique dont l'héritabilité estimée est d'environ 80%, cependant ils n'expliquent environ que 5% de la variance phénotypique malgré des études portant sur des dizaines de milliers d'individus. » (Manolio et *al.,* 2009, p. 747)

Ce problème de l'héritabilité manquante, au départ identifié pour des traits complexes comme la taille humaine, a été mis en évidence pour toutes les maladies qui avaient constitué un « succès » pour les associations pangénomiques, comme le montre la Figure 8. Dans le cas du diabète de type 2 par exemple, dont l'héritabilité est considérée comme relativement élevée, environ 95% des éléments qui expliquent la variance génétique sont manquants.

Maladie	Héritabilité	Nombre de locus trouvés	Fraction de l'héritabilité expliquée
Dégénérescence maculaire liée à l'âge	0,71	5	50 %
Maladie de Crohn		32	20 %
Lupus érythémateux disséminé		6	15 %
Diabète de type II	0,40	18	6 %
Cholestérol HDL		7	5,2 %
Taille	0,80	40	5 %
Infarctus du myocarde précoce		9	2,8 %

Figure 8 : Exemples de maladies pour lesquelles se pose le problème de l'héritabilité manquante (Jordan, 2010, p. 542). Pour toutes les maladies, le nombre de locus qui sont associés à une susceptibilité accrue à la maladie et qui ont été identifiés par des études d'associations pangénomiques (GWAS), est indiqué. Pour toutes les maladies, est également mentionnée la fraction de l'héritabilité expliquée par ces locus, par rapport à l'héritabilité estimée de la maladie dans la population sur laquelle a été conduite l'étude d'association pangénomique correspondante. Lorsqu'on dispose de valeurs d'héritabilité totale (une même valeur d'héritabilité est retrouvée systématiquement dans différentes populations), celles-ci sont mentionnées (2e colonne).

3.2.3. À la recherche de « l'héritabilité perdue »

Où est passée cette héritabilité manquante et quelles conséquences en tirer quant à l'intérêt des associations pangénomiques et à l'architecture allélique des maladies ? Plusieurs hypothèses ont été proposées pour faire sens de l'héritabilité perdue (Eichler et *al.,* 2010 ; Jordan, 2010 ; Lee, Wray, Goddard et Visscher, 2011 ; Manolio *et al.,* 2009*)*. Nous avons identifié deux grandes familles d'hypothèses : la première (l'existence d'une autre source de variation commune dans le génome) remet en cause le choix des marqueurs dans les associations pangénomiques, la seconde est beaucoup plus radicale parce qu'elle propose de repenser selon différents modèles (modèle infinitésimal, modèle des allèles rares, modèle de l'héritabilité au sens large) l'architecture génétique des maladies hors de l'hypothèse « maladie commune, variant commun » sur laquelle sont fondées les associations pangénomiques. Nous nous attarderons en particulier sur la troisième variante de cette deuxième famille d'hypothèses (le modèle de l'héritabilité au sens large) qui est susceptible de remettre profondément en cause la conception interactionniste quantitative.

3.2.3.1. Une autre source de variation commune

Lors du projet HapMap, 500 000 polymorphismes nucléotidiques simples ont été choisis comme marqueurs de référence, mais le consortium a aussi génotypé des millions de SNP et a tenté d'estimer l'importance de la variation génétique entre les humains. Selon le projet HapMap, si on prend les génomes de deux individus choisis au hasard, on ne trouvera pas plus de 0.1% de variation génétique et celle-ci a essentiellement la forme de polymorphismes nucléotidiques simples (The International HapMap Consortium, 2005). C'est ce qui explique le choix des SNP comme marqueur privilégié des associations pangénomiques. Ces estimations du projet HapMap ont cependant été remises en question par la découverte de la variabilité du nombre de copies des gènes – ou plutôt par sa redécouverte. On appelle « CNV » (Copy Number Variants) des duplications ou au contraire des délétions de petites régions du génome. Ces zones dupliquées, tripliquées ou au contraire délétées ont été pour la première fois mises en évidence il y a plus de soixante-dix ans (Bridges, 1936) chez la drosophile : la duplication du gène Bar semblait responsable d'une morphologie particulière des yeux de la drosophile. Malgré l'accumulation progressive de corrélations entre la présence de CNV et l'apparition de maladies, l'intérêt pour cette source de variation génétique de grande échelle est très récente (Beckmann *et al.,* 2007). La définition officielle des CNV est proposée en 2006 (Redon *et al.,* 2006) : c'est un segment d'ADN d'au moins un kilobase et dont le nombre de copies est variable en comparaison à un génome de référence. Un CNV peut être simple dans sa structure, comme une duplication simple ou peut impliquer des gains ou des pertes complexes de séquences homologues en de multiples sites du génome.

Lorsque le problème de l'héritabilité perdue s'est posé, c'est donc tout naturellement vers les CNV que les regards se sont tournés, surtout à partir du moment où des évaluations systématiques des variations de nombre de copies ont permis d'affirmer que les CNV affectaient un dixième du génome et qu'à la lumière de cette nouvelle estimation, les différences entre les génomes de deux individus pris au hasard n'était pas de 0.1% mais de 1%. La nouvelle hypothèse était donc la suivante : c'est parce que les associations génomiques utilisent les SNP comme marqueurs de la variation commune qu'elles ne peuvent détecter certaines sources de variations

génétiques et qu'elles ne parviennent pas à rendre compte de l'héritabilité des maladies complexes. Cette nouvelle hypothèse permettait d'ailleurs de ne pas remettre en cause l'hypothèse « maladie commune, variant commun » ; la source de la variation commune avait seulement été mal identifiée. Les fabricants de puces ont alors immédiatement modifié et adapté les puces pour permettre de détecter les CNV. Malheureusement une nouvelle étude pangénomique à la recherche des CNV, à l'initiative du Wellcome Trust Case Control, portant sur 19 000 individus et huit maladies majeures n'ont révélé que trois associations statistiquement significatives entre locus CNV et maladies, toutes les trois étant déjà connues (Craddock *et al.,* 2010).

3.2.3.2. D'autres hypothèses sur l'architecture allélique de la maladie

D'autres auteurs ont remis en cause l'hypothèse « maladie commune, variant commun », dont nous avons bien précisé plus haut qu'elle manquait par ailleurs de preuves empiriques. Il faut encore une fois revenir sur la nature des SNP – supposés être les polymorphismes les plus fréquents du génome humain – pour comprendre que l'échec des GWAS peut être imputable à une technologie qui n'est pas configurée pour détecter les variants rares :

> « Les puces à ADN les plus perfectionnées ne peuvent comporter que quelques millions de segments d'ADN différents à leur surface (ce qui est déjà une performance assez extraordinaire). Du coup, elles n'analysent dans l'ADN qui leur est appliqué qu'un nombre limité de locus, disons un million pour fixer les idées. Or il y a potentiellement trois milliards de points de variation possibles dans un génome haploïde humain. Les puces ciblent donc les polymorphismes les plus fréquents, les snip repérés par de nombreuses études et pour lesquels on s'est fixé arbitrairement une limite de fréquence de 5 % pour l'allèle mineur (l'allèle majeur étant alors présent à 95 % dans la population puisque ces locus sont quasiment tous bialléliques). On pouvait espérer que, comme formulé dans l'hypothèse *common disease / common variant*, la majeure partie du déterminisme génétique des maladies fréquentes serait portée par de tels

variants. Le résultat des milliers d'analyses GWAS menées à ce jour montre clairement que ce n'est pas le cas. » (Jordan, 2012, p. 893)

D'autres hypothèses ont été proposées pour expliquer l'architecture allélique des maladies (Gibson, 2009, 2012) :

(a) *Le modèle infinitésimal :* beaucoup de variants de petit effet : c'est l'hypothèse plus proche du modèle infinitésimal de Fisher. Dans cette hypothèse, à côté de quelques variants communs identifiés par les associations pangénomiques, la variabilité génétique des individus serait due à une multitude de variants dont l'effet mineur ne parvient pas à être détectée par les associations pangénomiques. L'héritabilité ne serait pas tant manquante que cachée – c'est la puissance d'identification d'allèles à effet mineur sur le risque de la maladie qui serait mise en cause.

(b) *Le modèle des allèles rares :* de nombreux allèles rares d'effet large : ce modèle a été proposée dès le début des années 2000, en réponse au modèle CVCD « common variant, common disease » (Pritchard, 2002) . Ce modèle postule que la variance génétique des maladies complexes serait majoritairement due à des variants rares, modérément pénétrants, dont la fréquence allélique serait typiquement inférieure à 1 %, et qui seraient apparus récemment dans la population humaine. Dans ce type de modèle, la schizophrénie par exemple, devrait être comprise comme une centaine de maladies distinctes mais similaires, chaque forme étant due à quelques variants rares. Comme ces variants sont rares et que les associations pangénomiques ne détectent que les variations communes, c'est du côté de ces allèles rares que se cacherait l'héritabilité manquante.

Il y a de nombreux arguments en faveur ou en défaveur de chacune de ces hypothèses (voir, en particulier, Gibson, 2012) mais la réalité est probablement moins tranchée, d'autant plus qu'un phénomène d'interactions synthétiques a été mis en

évidence qui complique encore notre compréhension de l'architecture allélique des maladies[17] (Dickson *et al.,* 2010).

En définitive, les connaissances scientifiques actuelles dessinent un portrait « mixte » de l'architecture allélique des maladies et soulignent le fait que certaines hypothèses biologiques ont été délaissées et/ou sont difficiles à tester. Ainsi, la Figure 9 représente les différentes formes d'architecture allélique des maladies en les classant en fonction de la fréquence allélique (en abscisse) et du risque conféré par la présence de l'allèle pathogène. La zone délimitée en diagonale par les pointillés indique le type de variants alléliques que la recherche en génétique a jusqu'ici privilégié. On a donc à l'extrême gauche de la diagonale des allèles rares responsables des maladies mendéliennes – ce sont eux qui ont fait les premiers succès de la génétique inverse. À l'autre extrémité de la diagonale, on trouve des variants communs impliqués par les associations pangénomiques dans les maladies communes ou complexes. Au milieu se trouve une catégorie intermédiaire qui peut être présente dans les maladies mendéliennes comme dans les maladies complexes et qui est bien symptomatique d'ailleurs du continuum allélique qui existe entre ces deux types de maladies : ce sont des variants peu fréquents, à effet intermédiaire. Enfin, à l'extérieur de la diagonale, se trouvent les variants que l'on a encore des difficultés à détecter – d'une part les variants rares à effet faible pour lesquels les associations pangénomiques ne sont pas adaptés et pour lesquels se posent des problèmes de puissance statistique, d'autre part des variants communs à effet fort pour quelques maladies communes.

[17] Ces travaux semblent indiquer que certains individus qui ont en commun un certain variant fréquent en raison de leur généalogie, peuvent par ailleurs par hasard, concentrer des variants rares à effet large. Ce phénomène, appelé « association synthétique », serait assez fréquent et fausserait potentiellement les résultats des associations pangénomiques. En effet, lorsque des associations pangénomiques sont recherchées sur une population de ce type, l'étude va détecter un variant fréquent conférant un risque faible à une fraction notable de la population, alors que seule une petite partie de cette population possède en fait un variant rare conférant un risque élevé. Si tel est bien le cas, cela signifie que certaines associations pangénomiques ont pu identifier à tort des variants communs à effet faible, alors que se cachaient des variants rares à effet large.

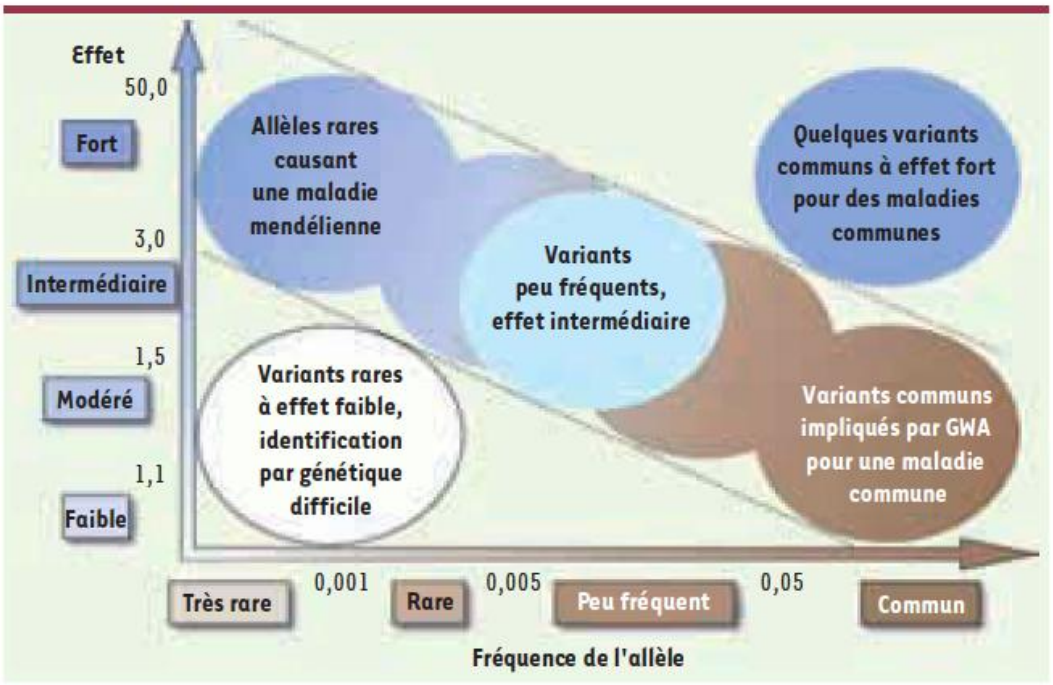


Figure 9 : Différents types d'architectures alléliques des maladies (Jordan, 2010, p. 542). Les différents types d'architectures alléliques des maladies sont classés selon la fréquence de l'allèle en abscisse et le risque relatif conféré par l'allèle pathogène en ordonnée. La diagonale transversale représente les variants sur lesquels la recherche en génétique médicale s'est concentrée jusqu'ici.

C'est ainsi tout naturellement que le retour de l'hypothèse des variants rares d'effet intermédiaire a entraîné le développement de nouvelles technologies, plus adaptées à la détection de ces variants que les associations pangénomiques, comme le séquençage « exome entier ». Un point de l'organisation du génome que nous n'avons pas abordé jusqu'ici est le fait que les séquences codantes ne représentent qu'une petite partie du génome, appelée « exome ». Ce sont pourtant logiquement ces régions codantes qui concentrent la majorité des mutations potentiellement responsables des maladies complexes. Pour augmenter la capacité de détection des mutations rares, une nouvelle méthode a donc été proposée (Choi *et al.,* 2009) consistant à séquencer l'ensemble de ces séquences codantes – ce qui permettrait en théorie d'identifier la totalité des variations de séquences codantes entre des patients malades et des patients sains pour une maladie donnée (Do, Kathiresan et Abecasis, 2012 ; Goh et Choi, 2012). Deux articles récents (Tennessen *et al*., 2012 ; Nelson *et al.*,

2012)[18] utilisant cette méthode de séquençage des exomes, sont venus renforcer l'hypothèse que les variants rares sont eux-mêmes assez abondants non pas au niveau de chaque génome individuel (qui reste principalement dominé par les SNP) mais au regard de la population générale. Les deux études vont dans le même sens, confirmant l'abondance des variants rares dans les génomes des deux populations étudiées mais surtout démontrant un effet important de ces variants dont l'altération aboutit fréquemment à une perte de fonction de l'allèle. Il est ainsi estimé que chaque personne porterait entre 30 et 100 substitutions aboutissant à une perte de fonction de l'allèle concerné (une vingtaine de ses substitutions intervenant à l'état homozygote). Il semble donc qu'une partie de l'héritabilité manquante puisse être expliquée par les variants rares – mais Bertrand Jordan met en garde contre les difficultés d'une mise en évidence de ces variants rares qui viennent à la fois des méthodes d'identification statistique de ces variants rares mais aussi de l'hétérogénéité génétique des différentes ethnies de population (Jordan, 2012).

3.2.3.3. Le modèle « de l'héritabilité au sens large »

Il reste un troisième modèle (c) que nous n'avons pas évoqué pour le moment et que Gibson nomme *« modèle de l'héritabilité au sens large »* qu'il définit ainsi :

> « Le modèle de l'héritabilité au sens large suppose que les contributions additives des variants communs et les larges effets des variants rares sont insuffisants pour expliquer l'héritabilité manquante. Les partisans de ce modèle renvoient à une longue histoire de détection d'interactions gène x gène (interactions GxG, aussi appelées épistasis) et d'interaction gène x environnement (interactions GxE) dans la recherche génétique quantitative sur les organismes modèles et notent le nombre croissant d'études documentant les effets épigénétiques, notablement les contributions génétiques d'empreinte parentale et les modes d'hérédité par méthylation de l'ADN. Comme les associations pangénomiques mesurent seulement l'effet moyen des allèles sur des milliers d'individus, elles seraient incapables de capturer

[18] (Pour un commentaire accessible et détaillé de ces deux articles aux méthodes statistiques complexes, on peut se reporter à Jordan, 2012)

> l'hétérogénéité de la taille des effets au niveau familial qui serait la marque de fabrique de ces éléments plus larges de l'architecture génétique. » (Gibson, 2012, p. 137)

Ce modèle mérite un développement conséquent et indépendant des deux modèles précédents pour au moins trois raisons distinctes. Premièrement, il est d'une tout autre nature que les deux modèles précédemment évoqués (modèle infinitésimal et modèle des variants rares). Le modèle infinitésimal et le modèle des variants rares s'opposent directement à l'hypothèse des variants communs – ce sont des hypothèses sur l'architecture allélique des maladies, des alternatives au modèle « variant commun, maladie commune ». Si on revient à la définition de l'héritabilité comme ratio de la variance génétique par la variance phénotypique, on comprend que ces deux hypothèses remettent en question le numérateur : il resterait d'autres variants alléliques rares à larges effets ou multiples à petits effets qui rendent compte de la variance génétique et qui n'ont pas été recherchés et donc pris en compte jusqu'ici. Le « modèle de l'héritabilité au sens large » porte les germes d'une critique beaucoup plus radicale, qui remet en cause la définition même de l'héritabilité et c'est sur ce point que nous allons nous concentrer dans les prochains paragraphes. Deuxièmement, le nom de ce modèle fait explicitement référence à la distinction entre l'héritabilité stricte et l'héritabilité au sens large que nous avions évoquée plutôt. Dans les deux cas, la variance phénotypique est décomposée comme l'addition de la variance génétique et de la variance environnementale. Dans l'héritabilité au sens strict, la variance génétique est approximée par la variance génétique additive (c'est-à-dire par l'addition des effets additifs des gènes), alors que dans l'héritabilité au sens large, la variance génétique est l'addition de la variance génétique additive, de la variance due aux effets de dominance et de la variance due aux effets des interactions entre les gènes, aussi appelée variance épistatique. On comprend donc en quoi le modèle de l'héritabilité au sens large consiste bien en un sens à revenir à la définition de l'héritabilité au sens large, puisqu'il s'agit de prendre en compte dans le modèle la variance due aux interactions gène x gène, qui n'est plus considérée comme négligeable. Cependant le nom « modèle d'héritabilité large » est aussi trompeur car

dans la définition de l'héritabilité large, les interactions gène x environnement n'apparaissent pas. Ce modèle est donc plus qu'une simple substitution de la définition de l'héritabilité au sens strict par l'héritabilité au sens large. Troisièmement – et c'est en partie une conséquence du point précédent, le « modèle de l'héritabilité large » comprend deux critiques qui n'ont pas tout à fait les mêmes implications – les interactions gène x gène posent des difficultés moins importantes pour le concept même d'héritabilité que les interactions gène x environnement. En effet, comme on va le voir, une certaine interprétation du problème des interactions gène x environnement remet en cause l'idée même d'une partition de la variance phénotypique entre deux facteurs indépendants que seraient la variance génétique et la variance environnementale.

3.3. Deux concepts d'interaction

3.3.1. Les interactions gène x gène ou l'épistasie

Pourquoi les interactions gène x gène peuvent-elles être une réponse partielle au problème de l'héritabilité perdue et en quoi cette réponse est-elle différente des précédentes ? L'hypothèse des *Copy Number Variants* et les modèles RAME et infinitésimal ont un point commun : ils ne remettent pas en question les estimations qui ont été faites de l'héritabilité des maladies – seulement la question de l'identification des sources de la variance génétique susceptibles d'expliquer l'héritabilité. Au contraire, avec la question des interactions gène x gène, ce n'est pas la façon de rendre compte de l'héritabilité que l'on remet en question mais l'estimation de l'héritabilité totale à expliquer. En effet, le fait de ne pas prendre en compte les interactions gène x gène aurait pour conséquence de surévaluer l'estimation de l'héritabilité et donc de sous-estimer la portion de l'héritabilité expliquée par les variants génétiques :

> « La génétique humaine a été hantée par le mystère de « l'héritabilité manquante » des traits complexes. Bien que les études aient découvert plus de mille deux cents variants associés aux maladies et aux traits communs, ces

> variants semblent typiquement expliquer seulement une part mineure de l'héritabilité. La proportion de l'héritabilité expliquée par un ensemble de variants est le ratio de (i) l'héritabilité due à ces variants (numérateur) estimée directement à partir des effets observés par (ii) l'héritabilité totale (dénominateur) inférée indirectement à partir des données sur une population. Le point de vue dominant est que l'explication de l'héritabilité manquante réside dans le numérateur – c'est-à-dire dans des variants qui n'ont pas encore été découverts. S'il est certain qu'il reste de nombreux variants à découvrir, nous montrons ici qu'une portion substantielle de l'héritabilité manquante pourrait venir d'une surestimation du dénominateur, créant « une héritabilité fantôme ». Précisément (i) les estimations de l'héritabilité totale assument implicitement que le trait n'implique pas d'interactions génétiques (épistasie) entre les loci (ii) cette assomption n'est pas justifiée parce que les modèles avec interactions sont également cohérents avec les données observées et (iii) dans de tels modèles, l'héritabilité totale pourrait être beaucoup plus petite et ainsi la proportion d'héritabilité expliquée beaucoup plus large. Par exemple, 80% de l'héritabilité manquante aujourd'hui pour la maladie de Crohn pourrait être due à des interactions génétiques, si la maladie implique des interactions parmi trois chemins métaboliques. En somme, l'héritabilité manquante ne correspond pas nécessairement directement à des variants manquants, parce que les estimations actuelles de l'héritabilité totale pourraient être gonflées de façon significative par les interactions génétiques. » (Zuk *et al.,* 2012, 1193)

Il faut noter que Zuk ne remet pas en cause la question des variants manquants – il n'y a pas d'incompatibilité entre les hypothèses que nous avons précédemment décrites et celle de « l'héritabilité fantôme ». En revanche, la nature des hypothèses est clairement différente. Dans les modèles précédents, on considère que la source de la variance génétique n'a pas été correctement identifiée et il faut prendre en compte d'autres sources de variation ou des variations plus rares. Le modèle de « l'héritabilité fantôme » propose au contraire de remettre en cause la façon dont l'héritabilité est estimée en prenant en compte les interactions gène x gène dans cette estimation.

Quelle est la réponse des chercheurs à ce problème et pourquoi les interactions statistiques ne sont-elles pas prises en compte dans le calcul de l'héritabilité ? Dans ce même article, Zuk approfondit l'exemple de la maladie de Crohn. Pour cette maladie inflammatoire chronique de l'intestin, les associations pangénomiques ont permis l'identification de 71 loci associés à une augmentation du risque de développer la maladie. Si on suppose, comme c'est le cas la plupart du temps, que les maladies ont une architecture génétique strictement additive, c'est-à-dire que chaque effet génétique est indépendant des autres effets génétiques, ces loci expliquent environ 21,5% de l'héritabilité totale estimée. Mais de nombreux auteurs ont souligné qu'il y a peu de raisons biologiques pour penser que les effets génétiques chez un individu donné soient additifs. C'est pourquoi Zuk propose d'introduire un modèle biologique très simple – la limitation des réactifs. Ce modèle consiste à dire que les processus biologiques dépendent souvent de la valeur limite des réactifs, de la même façon qu'une réaction chimique entre deux réactifs s'arrête si l'un des composants vient à manquer. En admettant que dans la maladie de Crohn trois réactifs soient concernés par un modèle de ce type, on passe d'une héritabilité expliquée de 21,5% à une héritabilité expliquée de 60%. Ce qui est en filigrane dans la démonstration de Zuk, c'est une distinction entre les interactions gène x gène telles qu'elles sont détectées statistiquement au niveau d'une population et leur interprétation biologique à l'échelle d'un individu. Or, cette distinction entre deux concepts d'interaction gène x gène (ou épistasie) est récurrente dans la littérature contemporaine. Les dénominations de ces deux concepts ne sont pas toujours les mêmes et ne recouvrent pas toujours les mêmes définitions. Ainsi, Philips a créé un vocabulaire spécifique pour distinguer entre trois types d'épistasie : l'épistasie fonctionnelle, l'épistasie compositionnelle et l'épistasie statistique (Philips, 2008). Il faudrait consacrer un travail à part entière à la traque des différents concepts d'épistasie – dans les limites de ce travail, nous nous reposerons sur la distinction adoptée par la majorité de la communauté scientifique entre « interaction statistique » et « interaction biologique » et qui est supposée remonter aux deux définitions de l'épistasie proposées par Bateson et Fischer.

William Bateson aurait été le premier à utiliser le terme « épistatique » en 1909 (Bateson, 1909), dans le contexte de l'explication des déviations de l'hérédité

mendélienne, afin de décrire l'effet masquant d'un allèle sur l'expression d'un autre allèle. Cordell donne un bon exemple d'un tel effet (Cordell, 2002). Dans la Figure 10, on s'intéresse à la couleur des poils de la souris, en admettant que deux loci (G et B), qui chacun ont deux allèles (*G* et *g* pour le locus G, *B* et *b* pour le locus B) déterminent cette couleur et que trois phénotypes (blanc, gris ou noir) peuvent résulter de la combinaison de ces allèles à ces deux loci sont possibles. Quels que soient les allèles présents au locus B, nous voyons que l'allèle *G* est dominant par rapport à l'allèle *g* (chaque fois que l'allèle *G* est présent, le phénotype est gris). En revanche, si on se place du point de vue du locus B, on voit que l'allèle *B* est dominant par rapport à l'allèle *b*, uniquement lorsque le génotype au locus G est *g/g*. En revanche, lorsque le génotype au locus G comporte au moins un allèle *G*, alors l'effet du locus B ne s'exprime pas et le phénotype est toujours gris. C'est dans cette situation que l'allèle *G* au locus G est dit « épistatique » par rapport à l'allèle *B* au locus B, l'allèle *B* au locus B étant alors dit « hypostatique ». Cette première définition des interactions épistatiques est considérée comme étant à l'origine du concept d'épistasie biologique parce qu'elle décrit l'effet de la substitution simultanée de deux allèles situés à deux loci différents sur le phénotype d'un individu. Par extension, on a pu parler d'épistasie biologique pour discuter de toutes les formes d'interaction entre gènes qui ont un effet sur le phénotype d'un individu, ce qui comprend entre autres les effets de régulation de réseaux de gènes entre eux.

	Genotype at locus G		
Genotype at locus B	*g/g*	*g/G*	*G/G*
b/b	White	Grey	Grey
b/B	Black	Grey	Grey
B/B	Black	Grey	Grey

Figure 10 : Représentation des phénotypes (exemple : couleur de pelage) obtenus pour deux allèles interagissant de façon épistatique selon la définition de Bateson (Cordell, 2002, p. 2463)

Le concept d'interaction statistique serait au contraire hérité de Fisher. Dans ce même article de 1918 dans lequel il analyse la décomposition de la variance

phénotypique entre variance génétique et variance environnementale (Fisher, 1918), il analyse les différentes composantes de la variance génétique. Après avoir posé l'hypothèse que la variance génétique est essentiellement additive, il reconnaît deux effets de déviation possible à ce modèle additif : d'abord l'effet de dominance, puis un effet qu'il nomme « épistasie » et qui est la déviation statistique entre les effets additifs attendus de deux facteurs mendéliens et les effets observés sur leur contribution quantitative à un phénotype donné :

> « Il y a dans la dominance une certaine latence. Nous pourrions dire que les effets somatiques de changements génétiques identiques ne sont pas additifs et que pour cette raison, la similitude génétique des relations est partiellement obscurcie dans l'agrégation statistique. Une déviation similaire de l'addition des effets surimposés peut survenir entre différents facteurs mendéliens. Nous pouvons utiliser le terme « épistasie » pour décrire une telle déviation, qui, bien que potentiellement plus compliquée, a des effets statistiques similaires à celui de dominance. » (Fisher, 1918, p. 404)

L'épistasie est donc définie par Fisher comme une déviation du modèle linéaire et additif des effets de différents allèles situés à différents loci sur leur contribution quantitative à un phénotype donné dans une population. C'est là une des grandes différences qui perdurent jusqu'à aujourd'hui dans l'opposition entre interaction statistique et interaction biologique. L'épistasie statistique est essentiellement l'épistasie détectée par les modèles mis en place pour identifier des variants génétiques causalement associés à la survenue ou à l'augmentation de risque d'une maladie. C'est un concept populationnel – ce qui est détecté c'est la façon dont les données s'écartent d'un modèle linéaire d'addition des effets génétiques, la façon dont l'effet combiné d'une substitution d'allèles en deux loci différents s'écarte de l'effet séparé de cette substitution et cette estimation se fait non au niveau d'un individu mais d'une population tout entière. C'est donc une moyenne des effets épistatiques ayant lieu en différents loci sur une population génétiquement hétérogène, le risque étant que certains effets épistatiques soient minorés ou au

contraire amplifiés par cette méthode de détection. L'épistasie biologique est au contraire un phénomène mécaniste mettant en jeu des interactions physiques entre les gènes et qui s'évalue au niveau d'un individu donné.

Dans le contexte du développement des associations pangénomiques et de la quête de l'héritabilité perdue, le problème est donc de comprendre la nature du lien entre épistasie statistique et épistasie biologique. En effet, on a reproché aux études d'association pangénomiques de tester l'association entre un variant donné et la maladie de façon isolée, sans prendre en compte le fait que des interactions entre les variants puissent avoir lieu. La réponse des généticiens quantitatifs a donc été de proposer un ensemble de modèles statistiques pouvant être potentiellement implémentés dans les associations pangénomiques et permettant de tester les interactions épistatiques (Cordell, 2009). Mais les partisans d'une conception biologique des interactions ne se limitent pas à pointer les insuffisances statistiques des modèles, ils proposent de repenser radicalement différemment le rapport entre interaction statistique et interactions biologiques. En effet, il faut d'abord garder à l'esprit que détecter une épistasie statistique c'est détecter une moyenne des effets épistatiques qui ont lieu entre différents allèles situés à différents loci dans une population génétiquement hétérogène – l'interaction statistique ainsi détectée ne revient donc pas à détecter une épistasie biologique mais plutôt à pointer une multitude d'effets biologiques sans qu'il soit forcément possible de savoir quels modèles biologiques sous-tendent l'effet statistique démontré et au risque de passer à côté de certains effets épistatiques biologiques qui sont trop petits pour pouvoir être détectés. Il n'est donc pas sûr que l'épistasie biologique puisse être détectée par les méthodes de l'épistasie statistique.

Moore et Williams (Moore et Williams, 2009, p. 312) ont ainsi le mérite de souligner les problèmes que pose la distinction entre les méthodes de détection statistique des interactions et leur interprétation biologique : ce n'est pas seulement que les modèles linéaires ont des difficultés méthodologiques inhérentes aux tests de multiples interactions (problème de puissance statistique, de taille d'échantillons), la difficulté est que les modèles linéaires sont construits pour amplifier les « effets principaux », c'est-à-dire les effets additifs des allèles individuels au détriment des interactions entre gènes. Ce problème de la possibilité de détecter statistiquement les

interactions biologiques se pose avec d'autant plus d'acuité pour les associations pangénomiques, qui testent, comme on l'a rappelé, des polymorphismes nucléotidiques surreprésentés dans une population malade par rapport à la population indemne de maladie. Il faut en effet garder à l'esprit que ces variants surreprésentés ne sont pas nécessairement les variants causalement associés à la maladie, mais simplement parfois les variants en déséquilibre de liaison avec ceux qui sont véritablement associés à la maladie :

> « Il peut sembler raisonnable de supposer que l'épistasie fonctionnelle sous la forme d'interactions entre des biomolécules ou d'interactions protéine-protéine est une composante ubiquitaire des chemins biologiques sous-jacents qui déterminent la progression des maladies. Cependant, cela ne signifie pas que cette épistasie sera détectée comme une interaction mathématique ou statistique, particulièrement si les variables qui sont examinées sont, comme dans beaucoup de cas, de simples substituts pour les vrais variants causaux sous-jacents qui sont corrélés avec les variants examinés en raison du déséquilibre de liaison ». (Cordell, 2009, p. 404)

On comprend à présent mieux l'intérêt de la méthode utilisée par Zuk : celui-ci n'a pas essayé de tester un modèle linéaire amélioré pour mieux détecter les interactions statistiques entre gènes dans la maladie de Crown, il a proposé d'utiliser un modèle biologique d'interactions gène x gène et de tester l'effet de ce modèle biologique sur la proportion d'héritabilité expliquée par les variants génétiques. Mais ce faisant, l'expérience de Peter Zuk pose d'autres questions : en admettant qu'on utilise des modèles biologiques pour tester les interactions entre gènes dans une maladie précise, quels modèles biologiques faut-il utiliser ?

> « Malheureusement, comme nous l'avons vu, il n'y a pas de correspondance précise entre les modèles biologiques d'épistasie et ceux qui sont motivés par une perspective statistique. Nous aimerions réaliser des tests statistiques et interpréter biologiquement le résultat, mais cela n'est en général pas possible.

> L'interaction statistique n'implique pas nécessairement l'interaction au niveau biologique et mécaniste. Une rapide revue de la littérature épidémiologique révèle les difficultés majeures qui existent à ingérer une signification biologique à partir de données quantitatives mesurant le risque de développer la maladie comme résultat. Le problème, c'est que tout ensemble de données et tout modèle statistique peuvent être généralement obtenus à partir d'un grand nombre de mécanismes et de modèles sous-jacents pour le développement de la maladie complètement différents les uns des autres. Par exemple, il a été montré que cinq mécanismes causaux très différents mènent tous à un modèle multiplicatif pour les données utilisées dans l'investigation des effets conjoints de deux facteurs de risque. C'est uniquement si un modèle biologique a priori peut être postulé en détail qu'il est probable que la modélisation statistique puisse éclairer les mécanismes biologiques sous-jacents. » (Cordell, 2002, p. 2465–2466)

Pour que l'épistasie statistique reflète ou corresponde à l'épistasie biologique et permette d'éclairer les mécanismes épistatiques dans le développement des maladies dont le caractère ubiquitaire a été souligné par de nombreux commentateurs (Moore, 2003), il ne suffit ni d'augmenter la puissance de détection statistique des interactions épistatiques dans les modèles linéaires ni d'améliorer les modèles linéaires en leur permettant de détecter des interactions épistatiques non linéaires et de ne pas conditionner la détection des interactions épistatiques à la détection préalable d'effets alléliques individuels, ni même d'utiliser des modèles biologiquement plausibles et non linéaires pour détecter ces interactions. Pour que l'épistasie statistique reflète l'épistasie biologique, il faut renverser le postulat hérité des associations pangénomiques selon lequel l'agnosticisme biologique est la meilleure des solutions et développer pour elle-même l'étude des modèles biologiques d'interactions épistatiques. C'est en utilisant cette connaissance biologique préalable que la détection statistique de l'épistasie pourra avoir une interprétation biologique. Nous allons à présent nous tourner vers le problème des interactions gène x environnement et voir dans quelle mesure elles posent des problèmes similaires mais aussi des problèmes additionnels à ceux soulevés par les interactions gène x gène.

3.3.2. Les interactions gène x environnement

3.3.2.1. Deux concepts d'interaction gène x environnement

En quoi les associations pangénomiques ont-elles délaissé les interactions gène x environnement et en quoi ces dernières expliquent-elles en partie la question de l'héritabilité perdue ? Pour bien le comprendre, il est utile de reprendre l'histoire du concept d'interactions gène x environnement – ou plutôt, comme l'ont affirmé Paul Griffiths et James Tabery (Griffiths et Tabery, 2008 ; Tabery, 2007, 2009) dans une série d'articles, l'histoire des deux concepts d'interaction gène x environnement que Tabery appelle concept « biométrique » d'interaction gène x environnement (concept GxE_B) et concept « développemental » d'interaction gène x environnement (concept GxE_D). L'origine du concept biométrique remonte encore une fois à Fisher. Dans l'article de 1918, le concept d'interaction gène x environnement n'apparaît pas. Selon Tabery, c'est bien plus tard, dans les années vingt, en travaillant sur les relations entre la culture de différentes variétés de pommes de terre et différents types d'engrais que Fisher soulève le problème des interactions gène x environnement, c'est-à-dire en l'occurrence la possibilité que deux variétés de pommes de terre puissent réagir de façon différente à différentes conditions environnementales. Pour tester si une telle interaction existait, Fisher considère que les différences de croissance que l'on observe dans les différentes variétés de pommes de terre est due à deux variables : la variété elle-même et le traitement à l'engrais choisis. Il mesure alors le rendement moyen de chaque variété de pommes de terre, indépendamment du type de traitement reçu et le rendement moyen pour chaque traitement appliqué indépendamment de la variété cultivée. Puis il compare ces rendements moyens avec le rendement moyen de la totalité de la culture. À partir de ces résultats, il calcule la variation due au traitement, la variation due à la variété et ce qu'il appelle la « déviation de la formule de la somme » qui correspond à l'interaction entre les différents engrais et les différentes variétés. Il calcule alors que cette « déviation de la formule de la somme » existe mais n'est pas statistiquement significative, c'est-à-dire qu'elle n'est pas suffisamment importante pour ne pas être due au hasard. Tabery en conclut que la question des interactions gène x environnement est bien prise en compte dans le modèle de Fisher

mais qu'elle est considérée comme négligeable, une complication possible dans la partition de la variance environnementale mais dont il n'est a priori pas nécessaire de s'inquiéter puisque, si de telles interactions existaient, on pourrait les détecter statistiquement, c'est-à-dire qu'il y aurait une déviation significative du modèle additif de la partition de la variance phénotypique entre variance génétique et variance environnementale (Tabery, 2007, p. 964).

Cependant, dans les années trente, un autre concept d'interaction GxE, le concept développemental est apparu sous la plume de Lancelot Hogben qui a même engagé une correspondance avec Ronald Fisher à ce sujet, en utilisant les résultats obtenus par Krafkan, en particulier un graphique représentant la norme de réaction du développement du nombre de facettes des yeux de deux populations génétiques différentes de drosophile en fonction de la température. Hogben insiste sur la difficulté à interpréter les concepts de variance génétique et de variance environnementale quand deux génotypes différents réagissent différemment à la même variété d'environnements. Nous ne pouvons pas reprendre ici la totalité de la démonstration de Tabery – mais il apparaît clairement dans leur correspondance que Fisher et Hogben ne parlent pas du même phénomène : tandis que Fisher considère les interactions GxE comme un phénomène probablement rare, détectable statistiquement et qu'il définit comme la mesure de la déviation du modèle additif de la variance phénotypique, Hogben considère les interactions gène x environnement comme une troisième source de variabilité au sein des populations, comme un phénomène lié au développement de l'organisme et qui d'ailleurs devrait être exploré non pas par une mesure statistique mais par l'embryologie expérimentale. C'est ce concept d'interaction gène x environnement que Tabery appelle « développemental », GxE_D.

James Tabery retrace ensuite la persistance de ces deux concepts d'interaction tout au long de l'histoire de la biologie, en revenant entre autres sur l'opposition entre Lewontin et Jensen à ce sujet, que nous avons évoquée plus tôt dans ce travail. Alors que Lewontin insiste sur le fait que l'héritabilité est toujours une analyse de la causalité locale d'une population dans un environnement donné et que l'on ne peut extrapoler les résultats de cette analyse locale dans d'autres environnements en raison des interactions gène x environnement qui ne permettent pas de prédire ce qui se

passerait dans un environnement non testé, Jensen s'est défendu en reprochant à Lewontin et aux critiques de l'héritabilité de confondre le concept d'interaction statistique détectable dans l'analyse de la variance des populations avec les interactions entre les gènes et l'environnement qui existe au niveau du développement d'un individu.

Mais alors quels rapports entretiennent les deux concepts d'interaction ? C'est peut-être le point le plus intéressant de l'analyse de Tabery, qui souligne que la controverse particulièrement violente entre Lewontin et Jensen a abouti à dissocier de façon radicale les deux concepts d'interactions. Lewontin conclut ainsi dans son article de 1974 que l'analyse de variances n'est d'aucune utilité pour comprendre les interactions qui ont lieu au cours du développement d'un phénotype entre génotype et environnement et qu'il faut au contraire se concentrer sur les normes de réaction et sur la compréhension des mécanismes du développement qui sous-tendent les interactions entre gène et environnement. Jensen, de son côté, a dissocié radicalement les objectifs de la génétique quantitative, capable de détecter statistiquement les interactions ayant lieu entre les gènes et un environnement donné de ce qu'il appelle l'« l'interactionnisme » des biologistes du développement qui affirmeraient que les interactions entre les gènes et l'environnement étaient à ce point liés dans le développement d'un individu qu'il était impossible de distinguer leurs contributions relatives :

> « La nature acrimonieuse de la controverse sur le QI a laissé une impression durable sur la génétique comportementale. Que l'on soit ou pas d'accord avec les conclusion de Jensen sur l'héritabilité du QI, la distinction entre le développement individuel et les différences entre individus (souvent décrits en termes de différents « niveaux d'analyse »), a mené à une forme de pluralisme isolationniste entre les généticiens quantitatifs du comportement et leurs critiques. L'idée était que le débat pouvait être désamorcé en isolant les généticiens quantitatifs du comportement intéressés par les questions de l'importance[19] des causes de la variation responsables des différences individuelles au niveau d'une population, des biologistes à l'esprit causal-

[19] Dans la version originale : *« How much ? questions »*

> mécaniste qui cherchent à savoir comment[20] les mécanismes causaux sont responsables du développement individuel au niveau de l'individu. » (Tabery, 2007, p. 970)

Cependant, comme le remarque Tabery, une autre voix s'est fait entendre plus récemment qui remet en cause ce « pluralisme isolationniste » entre biologistes de la génétique quantitative et biologiste du développement. Ainsi, des biologistes, comme Rutter ou Greenland, reprennent la distinction faite par Tabery entre un concept biométrique et un concept développemental d'interactions gène x environnement, en les distinguant sous les termes respectifs d'interaction « statistique » et d'interaction « biologique ». Mais au lieu de distinguer radicalement ces concepts d'interaction statistique et d'interaction biologique, ils en donneraient au contraire, selon Tabery, une vision complémentaire :

> « Bien que ces notions d'interaction soient distinctes pour Rutter, elles ne sont pas isolées l'une de l'autre comme des niveaux différents d'analyse. L'identification d'une interaction statistique pointe plutôt la direction pour rechercher l'interaction mécaniste sous-jacente. « L'identification d'une interaction statistique n'est rien de plus que l'indicateur qu'il y a une question de processus qui doit être prise en compte : c'est cette caractéristique qui donne à une telle interaction son importance à la fois théorique et pratique[21]. » » (Tabery, 2007, p. 971)

Nous avons repris l'argumentation de Tabery parce qu'elle nous paraît éclairer de façon très pertinente les débats qui ont eu lieu sur le concept d'interactions gène x environnement et sur le rôle que ce concept joue dans la remise en question du modèle additif de l'héritabilité. Le parallèle est d'ailleurs frappant avec ce que nous avons dit du débat sur les deux concepts d'interactions gène x gène et il y a fort à parier qu'une analyse historique détaillée de l'histoire de ce concept pourrait mettre

[20] Dans la version originale : *« How ? questions »*
[21] Tabery cite ici un article de Rutter et Pickles (Rutter et Pickles, 1991, p. 106)

en évidence la même distinction originelle entre un concept biométrique et un concept développemental d'interactions gène x environnement. Il nous semble cependant qu'on peut aller encore un peu plus loin dans l'analyse des rapports entre le concept statistique et le concept biologique des interactions gène x environnement. Pour cela, il faut dire quelques mots de l'actualité des recherches sur les interactions gène x environnement.

3.3.2.2. Les difficultés inhérentes à la détection des interactions gène x environnement et les problèmes posés par les associations pangénomiques

Comme nous l'avons montré, le concept d'interaction gène x environnement est connu depuis longtemps. Il n'est pas question dans le cadre de ce travail de décrire la totalité des méthodes statistiques employées à la détection des interactions gène x environnement – aussi ne ferons-nous que mentionner la longue tradition de recherche des interactions gène x environnement à travers les études de jumeaux monozygotes ou dizygotes et à travers les méthodes d'analyse de liaison pour nous concentrer sur la période qui débute au milieu des années 1990 et où a commencé à se développer l'approche gène-candidat pour étudier les interactions gène x environnement.

Nous avons déjà évoqué les difficultés que rencontrait l'approche gène-candidat quand il s'agissait de mettre en évidence un effet génétique pour une maladie complexe. Détecter une interaction gène x environnement pose le même type de problèmes (nécessité d'avoir des connaissances physiopathologiques sur la maladie, nécessité d'échantillons importants et à l'époque coût important des méthodes de séquençage) à quoi il faut ajouter trois difficultés supplémentaires qui concernent toute étude d'un facteur environnemental : l'évaluation avec précision de l'exposition environnementale, une taille d'échantillon encore plus importante pour assurer une bonne puissance statistique et le problème de l'hétérogénéité (Hunter, 2005) .

Pour développer ces trois difficultés, nous avons choisi de prendre l'article de 2002 de Caspi et Moffitt, article qui a connu un immense succès car c'est la première

fois qu'est mise en évidence une interaction gène x environnement dans une maladie psychiatrique. Dans leur article de 2002, Caspi et Moffitt se sont intéressés à deux populations génétiquement différentes, l'une étant porteuse d'un variant allélique du gène MAOA (monoamine oxidase A) codant une activité basse de la monoamine oxidase A, tandis que l'autre était porteuse d'un variant allélique du gène MAOA codant une activité élevée de la monoamine oxidase A. En effet, la MAOA est une enzyme qui inactive certains neurotransmetteurs et une déficience de la MAOA avait été associée au développement d'une personnalité antisociale, sans que ces études soient formellement répliquées. L'expérience a donc consisté à comparer comment se comportent ces deux populations (en termes d'activité antisociale), selon que les individus aient été exposés ou non dans leur enfance à de la maltraitance familiale. La maltraitance familiale comprenait des abus physiques et psychologiques, peu d'interactions entre la mère et l'enfant, la négligence, etc., et pouvait être considérée dans l'étude comme nulle, probable ou sévère.

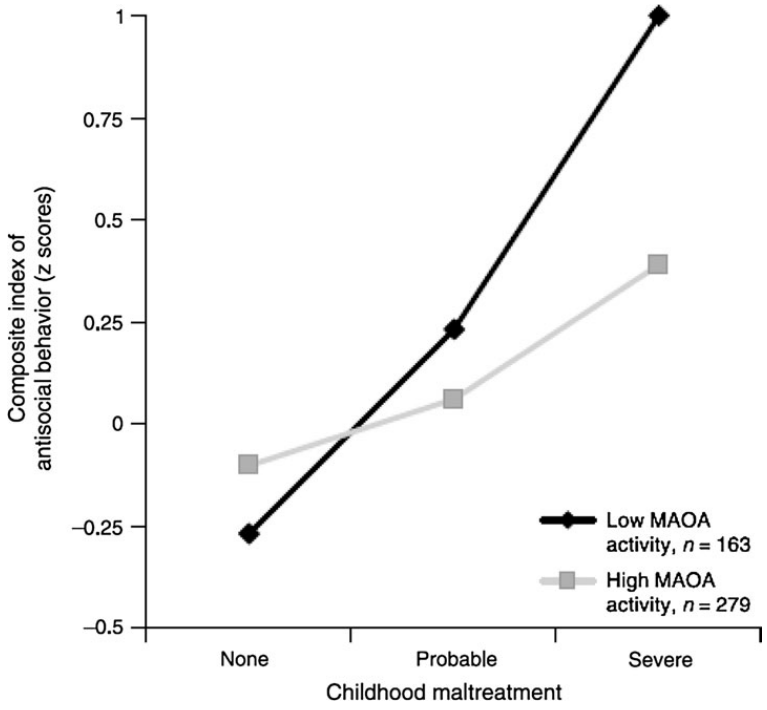


Figure 11 : Norme de réaction pour l'activité de la MAOA, la maltraitance dans l'enfance et le comportement antisocial (Source Caspi et Moffitt 2002)

Ce que montre le graphique ci-dessus (Figure 11), c'est que les individus qui ont une activité élevée de MAOA et qui selon les études précédentes sont le moins

disposés à développer des comportements antisociaux, ont un comportement qui varie en fonction de l'environnement auquel ils sont exposés : alors qu'ils ont plus de chance de développer des comportements antisociaux que les individus avec une faible activité de MAOA dans un même environnement dépourvu de maltraitance, la probabilité qu'ils développent un comportement antisocial augmente graduellement au fur et à mesure qu'ils sont dans un environnement avec une maltraitance de plus en plus sévère. Au contraire, les individus qui ont une faible activité de MAOA ont moins de chance de développer une activité antisociale dans un environnement sans maltraitance, mais leur probabilité de développer une activité antisociale augmente drastiquement au fur et à mesure que l'exposition à la maltraitance dans l'enfance a été sévère.

Précisément parce qu'ils se sont intéressés à une maladie psychiatrique, les travaux de Caspi et Moffitt ont soulevé au moins autant d'enthousiasme que de critiques et exemplifient fort bien les trois sources de difficultés supplémentaires que nous avons relevées. Les problèmes que pose l'exposition environnementale sont assez évidents dans ce cas précis : comment établir une échelle de maltraitance ? En l'occurrence, les auteurs de l'étude ont considéré comme des actes de maltraitance aussi bien la négligence que les abus physiques et psychologiques, est-ce légitime ? Si au lieu d'utiliser une échelle de trois catégories de maltraitance (absence de maltraitance, maltraitance probable, maltraitance sévère), on avait utilisé une échelle graduelle de la maltraitance de 0 à 10, aurait-on obtenu les mêmes résultats ? Y a-t-il une période critique pendant laquelle la maltraitance a un effet majeur sur les différents génotypes – par exemple très tôt dans l'enfance ou au contraire dans l'adolescence ? Pour donner une idée du suivi qui a été nécessaire en l'occurrence, les 422 individus masculins[22] de l'étude ont été génotypés à leur naissance puis suivis jusqu'à l'âge de 26 ans pour documenter leur histoire familiale. La deuxième catégorie de problèmes concerne la puissance statistique et la taille des échantillons nécessaires.

[22]Les individus faisaient partie d'une cohorte plus large étudiée dans le cadre des études multidisciplinaires de santé et de développement de Dunedin. Seuls les individus masculins ont été suivis car il est beaucoup plus simple chez les individus masculins de distinguer génétiquement une population à basse activité MAOA et une population à haute activité MAOA. En effet, le gène MAOA est situé sur le chromosome X : chez les femmes, le fait d'avoir deux chromosomes X oblige à distinguer entre trois groupes : un groupe porteur des deux allèles de haute activité MAOA, un groupe porteur des deux allèles de basse activité MAOA et un troisième groupe hétérogène qui ne peut pas être déterminé avec certitude car il n'est pas possible de savoir quel chromosome a été inactivé.

En l'occurrence, les auteurs ont suivi 442 individus masculins, dont 163 avec une faible activité de l'enzyme MAOA et 279 avec une forte activité de l'enzyme MAOA. En général, on estime que pour les études d'interactions gène x environnement avec approche gène-candidat, il faut compter des échantillons de quelques centaines de personnes. Enfin, le troisième type de difficulté concerne les problèmes d'hétérogénéité et de réplication. Plusieurs études (Haberstick *et al.,* 2005 ; Young *et al.,* 2006) ne sont ainsi pas parvenues à reproduire les résultats de Caspi et Moffitt, même si une méta-analyse semble avoir confirmé leurs résultats (Kim-Cohen *et al.,* 2006). Les problèmes de réplication sont monnaie courante dans tous les designs expérimentaux que nous avons passés en revue, mais la détection des interactions gène x environnement impose parfois de comparer des outils différents d'évaluation de l'exposition environnementale (par exemple des échelles différentes de maltraitance) qui ont des distributions ou des caractéristiques différentes d'exposition ou qui sont liés à des facteurs confondants différents – toutes choses qui augmentent le risque d'hétérogénéité dans les études et les difficultés de réplications.

De nombreuses raisons expliquent que les premières associations pangénomiques aient laissé de côté la question des interactions gène x environnement, en particulier si on reprend les difficultés que nous avions évoquées précédemment. En effet, dans les associations pangénomiques, l'approche est agnostique du point de vue biologique, il n'y a donc en général pas ou peu de connaissances sur la physiopathologie de la maladie concernée et sur d'éventuelles associations épidémiologiques entre certains facteurs environnementaux et certains variants génétiques. Or tester pour chaque SNP la possibilité d'interactions gène x environnement multiplie considérablement le nombre d'opérations statistiques déjà conséquent à réaliser dans une association pangénomique. Cela augmente aussi la taille des échantillons à considérer pour avoir une puissance statistique significative et les difficultés d'évaluation précise de l'exposition environnementale des individus concernés. Pour toutes ces différentes raisons, le problème des interactions gène x environnement n'a pas été abordé tant que les associations pangénomiques étaient dans leur première phase de succès et ce n'est que lorsque le problème de l'héritabilité perdue s'est posé avec acuité qu'on a considéré qu'il était nécessaire de s'en préoccuper.

3.3.2.3. Les interactions gène x environnement remettent-elles en cause le concept d'héritabilité ?

Comme dans le cas du problème des interactions gène x gène, on peut distinguer deux attitudes face au problème de la détection statistique des interactions gène x environnement. La première attitude consiste à proposer de nouvelles méthodes statistiques susceptibles d'amplifier la détection statistique des interactions gène x environnement ou de les rendre plus faciles. C'est ainsi qu'ont émergé récemment d'une part des revues de méthodologie pour mesurer les interactions gène x environnement dans les associations pangénomiques (Dempfle *et al.*, 2008 ; Thomas, 2010) et d'autre part des propositions de nouvelles méthodes comme celle des associations pangénomiques panenvironnementales[23] (Aschard *et al.,* 2012 ; van IJzendoorn *et al.,* 2011 ; Rakyan *et al.*, 2011). Pour ces auteurs, les interactions gène x environnement posent certainement un problème de détection statistique et nombreux sont ceux qui mettent en garde sur la difficulté d'interpréter biologiquement les interactions GxE détectées statistiquement. Ainsi, la recommandation de ne chercher à détecter statistiquement des interactions gène x environnement que lorsqu'un mécanisme biologique plausible a été identifié revient fréquemment :

> « Tout d'abord, il faut distinguer les études en fonction de leurs objectifs, selon qu'elles soient avant tout exploratoires (c'est-à-dire générant des hypothèses) ou confirmatoires (c'est-à-dire testant des hypothèses). […] Deuxièmement, une étude confirmatoire bien pensée des interactions GxE doit être fondée sur une hypothèse a priori justifiable d'une interaction entre un gène connu et un risque environnemental connu, liée d'une façon ou d'une autre à la fonction du gène, pour laquelle un mécanisme raisonnable d'interaction biologique existe. Seulement les hypothèses déjà spécifiées (c'est-à-dire avant la collection des données) et des tests statistiques peuvent être interprétés comme confirmatoires. Idéalement, il y a déjà des preuves d'une interaction entre l'exposition et les facteurs génétiques venant d'études génétiques

[23] C'est ainsi que nous traduisons faute de mieux l'expression *« Genome-Wide Environmental Studies »*.

> formelles (par exemple d'études sur les jumeaux ou d'analyses de ségrégation). » (Dempfle *et al.,* 2008, p. 1170)

Pour résumer, les auteurs de ces différents articles abordent avec prudence la question des correspondances entre interaction statistique et interaction biologique, mais leur point de focalisation premier est l'amélioration des méthodes de détection statistique, sans que la question des interactions gène x environnement ne vienne remettre en question le modèle de l'héritabilité.

La deuxième attitude est plus radicale, à plusieurs titres. Tout d'abord, tout en reconnaissant qu'il est important d'améliorer les moyens de détection statistique des interactions gène x environnement, elle consiste à admettre que ces améliorations auront des limites et que l'absence de détection d'une interaction statistique ne pourra presque jamais infirmer l'existence d'une interaction biologique (Uher, 2008, p. 23). Ces auteurs ont ainsi tendance à insister sur l'importance des validations expérimentales des interactions gène x environnement plutôt que sur leur détection statistique (Uher, 2008, p. 21). Deuxièmement, cette attitude, à l'instar de celle de Hogben, se traduit par une vision des interactions gène x environnement qui n'est pas simplement une déviation du modèle additif des effets majeurs des gènes et des effets majeurs de l'environnement, mais un fait ubiquitaire. Les interactions gène x environnement ne seraient pas un phénomène minoritaire mais au contraire un mécanisme décisif en physiologie et en physiopathologie (Uher, 2008, p. 14). En fait, selon les tenants les plus radicaux d'une telle thèse, il est évident d'un point de vue biologique que les interactions gène x environnement sont bien la règle et non l'exception et parler d'effet génétique indépendant ou d'effet environnemental indépendant n'a pas de sens.

> « La focalisation sur « les contributions relatives » du gène et de l'environnement comme des forces indépendantes est le cheval de bataille des généticiens quantitatifs du comportement et suppose que de telles forces peuvent agir indépendamment l'une de l'autre. En suivant la tradition biométrique de Fisher et Haldane, la variance dans l'expression d'un trait

> phénotypique dans une population est examiné en utilisant les modèles ANOVA qui partitionnent la variance phénotypique en fonction des facteurs génétiques et environnementaux. Il faut noter que dans de telles analyses, les influences « génétiques » sont définies statistiquement et non biologiquement et il en va de même bien sûr pour la définition des prétendues influences environnementales. […] Ainsi, les études de l'héritabilité des traits sélectionnés fournissent des preuves statistiques et non biologiques de l'influence génétique. Ce qui est irritant c'est l'interprétation de tels résultats et en particulier la notion « d'effets principaux » qui implique que le génome et l'environnement peuvent opérer indépendamment. Et bien sûr, nous devrions être tout aussi irrités par les effets principaux environnementaux que par ceux associés au génotype. A nouveau, le problème c'est l'interprétation et l'appréciation de ce que de tels résultats signifient réellement ou ne signifient pas. » (Meaney, 2010, p. 44)

Jusqu'ici on semble retrouver finalement les mêmes oppositions que celles évoquées par Tabery entre les partisans d'une conception statistique ou biométrique des interactions GxE et ceux d'une conception causale mécaniste ou développementale des interactions GxE et la même stratégie de « pluralisme isolationniste » – du moins de la part des partisans du concept développemental de GxE. Mais, et c'est sans doute là le point le plus décisif, la compréhension des mécanismes biologiques des interactions a beaucoup progressé aujourd'hui – ces mécanismes sont en partie ceux de l'épigénétique. Qu'est-ce que l'épigénétique ? Le concept a connu de multiples définitions (Bird, 2007 ; Jablonka et Lamb, 2002 ; Morange, 2002, 2005, 2011), et les limites de ce travail ne nous permettent certainement pas de reprendre en détail l'histoire de ce concept. Nous nous contenterons donc d'utiliser ici la définition récente proposée par Adrian Bird, parce qu'elle a vocation à unifier différents usages du terme épigénétique tout en envisageant les découvertes scientifiques les plus récentes comme les modifications épigénétiques transgénérationnelles :

> « La définition suivante pourrait unifier les événements épigénétiques : l'adaptation structurelle de régions chromosomiques afin d'enregistrer, de signaler ou de perpétuer des états d'activité modifiés. Cette définition est inclusive des marques chromosomiques, parce que les modifications transitoires associées à la fois à la réparation de l'ADN ou aux phases du cycle cellulaire, comme les changements stables maintenus à travers plusieurs générations cellulaires en font partie. Elle se focalise sur les chromosomes et les gènes en excluant implicitement la matrice architecturale tridimensionnelle des systèmes de membrane et des prions, excepté quand celle-ci modifie la fonction du chromosome. Est aussi incluse l'excitante possibilité que les processus épigénétiques soient les tampons de la variation génétique, qui reposerait sur un changement épigénétique (ou mutationnel) menant une combinaison identique de gènes à produire des résultats développementaux différents. » (Bird, 2007, p. 398)

Ce qui nous intéresse ici et qui intéresse les tenants de cette position radicale en faveur de l'importance des interactions biologiques, c'est la conjonction de deux facteurs : 1) les interactions biologiques gène x environnement sont en partie le reflet de mécanismes épigénétiques, 2) il est admis que certaines modifications épigénétiques peuvent être transmises des parents aux enfants et donc héritées sur une voire plusieurs générations. On comprend alors en quoi la question des interactions gène x environnement joue un rôle majeur dans la question de l'héritabilité manquante et peut être susceptible de modifier en profondeur le concept d'héritabilité. En effet, la critique de l'héritabilité portée par Lewontin revenait à réécrire l'équation de Fisher de la manière suivante :

$V_P = V_G + V_E + V_{GxE}$

en considérant, contrairement aux tenants de la tradition biométrique, que la variance due aux interactions gène x environnement (V_{GxE}) était loin d'être un phénomène statistique mineur mais peut-être au contraire un effet majeur. Mais le fait que les interactions gène x environnement puissent être partiellement héritées déplace le problème vers la prise en compte de l'hérédité non-génétique. Il faudrait sans doute écrire quelque chose qui se rapproche de ce que Etienne Danchin exprime

dans son modèle de l'héritabilité inclusive (voir Figure 12) : a) la variance phénotypique n'est pas l'addition mais le produit de la variance transmise et de la variance non transmise, b) la variance transmise est elle-même le produit de la variation génétique héritée et de la variation transmise non-génétique, c) la variance transmise non génétique peut elle-même se décomposer en différents composants comprenant entre autres la variance épigénétique héritée. Nous ne discutons pas des autres formes de variance incluses par Danchin dans son schéma et qui pourraient être sujets à débat, en particulier en ce qui concerne la variation sociale transmise. Nous voulons simplement utiliser ce schéma pour montrer que la deuxième position que nous avons identifiée et qui met l'accent avant tout sur les mécanismes biologiques des interactions gène x environnement n'est pas une simple redite des critiques de Lewontin, mais constitue bien une attaque nouvelle et différente du modèle de l'héritabilité.

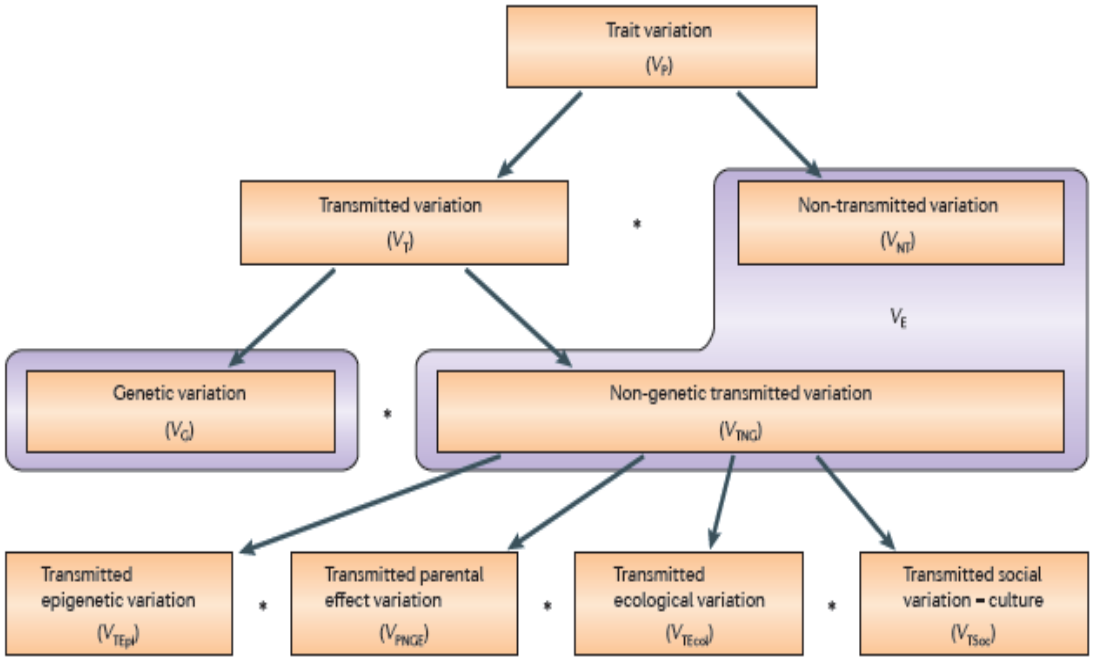


Figure 12 : La contribution quantitative des composants de l'héritabilité inclusive (Danchin *et al.*, 2011). Danchin propose une autre décomposition de la variance phénotypique d'un trait qui est le produit (et non l'addition) d'une variance transmise et d'une variance non transmise. Celles-ci peuvent à leur tour être décomposées plus finement.

On comprend alors mieux la cohérence et l'importance des critiques de cette deuxième attitude que Michel Meanery exprime ainsi :

> « Une indication d'héritabilité élevée pour un trait spécifique amène logiquement à la question de savoir si les variations génomiques au niveau de la séquence peuvent servir de mécanisme pour le passage des différences individuelles phénotypiques des parents aux enfants […]. Mais l'idée que les estimations de l'héritabilité sont l'équivalent d'un effet génétique est problématique. D'abord, de telles estimations ignorent les interactions gène x environnement. En effet, l'héritabilité équivaut à un effet principal « génétique » si et seulement si on assume qu'il n'y a pas de contribution gène x environnement. Deuxièmement, l'estimation de l'héritabilité reflète l'importance des mécanismes génomiques seulement si l'on assume qu'il n'y a pas d'autres mécanismes biologiques pour l'hérédité. Ceci est simplement faux. Il y a de multiples mécanismes potentiels d'hérédité, impliquant par exemple le passage de marques épigénétiques […] à travers la lignée germinale, le passage de molécules d'ARN maternel à l'embryon, le passage potentiel de protéines du prion des parents aux enfants, l'état biochimique des gamètes au moment de la conception et la transmission de nutriments, de bactéries ou d'anticorps depuis la circulation maternelle à celle des ses enfants, et ainsi de suite […]. Les variations dans la séquence génomique ne sont tout simplement pas le seul mécanisme de l'hérédité. Tous ces facteurs peuvent influencer et influencent le phénotype des descendants. C'est un sérieux problème que d'assumer que l'héritabilité équivaut aux effets principaux de la variation génomique. » (Meaney, 2010, p. 44)

Ajoutons enfin que cette remise en cause de l'héritabilité sur la mise en évidence de l'hérédité épigénétique est d'autant plus efficace que de nombreuses revues de littérature ont souligné l'importance croissante des modifications épigénétiques dans le développement de la maladie, accréditant ainsi par association[24] l'idée selon laquelle les interactions gène x environnement sont un phénomène ubiquitaire de la maladie (Feinberg, 2007 ; Jiang, Bressler et Beaudet, 2004 ; Petronis, 2010 ; Skinner, Manikkam et Guerrero-Bosangna, 2010).

[24] Par association, car il faut bien se souvenir que l'épigénétique ne désigne pas que les modifications chromosomiques stables mais aussi les modifications chromosomiques transitoires.

En résumé, selon la façon dont on aborde le concept d'interaction gène x environnement dans le problème de l'héritabilité perdue, on aboutit soit à une remise en question « modérée » du modèle de l'héritabilité (le modèle de l'héritabilité en lui-même n'est pas remis en cause mais les interactions gène x environnement font partie des éléments qu'il convient de prendre en compte pour régler le problème de l'héritabilité perdue), soit à une remise en question beaucoup plus profonde qui questionne la partition même de la variance phénotypique entre variance environnementale et variance génétique.

3.3.3. Que nous apprend le cadre interactionniste de la génétique quantitative ?

Comme le dit James Tabery (Tabery, 2009a), la philosophie de la biologie a été marquée par l'héritage de Lewontin, au point que nombreux sont ceux qui partagent ses critiques les plus acerbes du concept d'héritabilité et que d'autres, dépassant très largement les intentions initiales de Lewontin, vont jusqu'à dénier tout intérêt à ce concept. Le concept d'héritabilité est pourtant bien vivant : il a joué et joue toujours un rôle majeur dans la génétique des maladies. C'est pourquoi il nous a semblé essentiel de revenir en détail sur la forme d'interactionnisme que sous-tend un tel concept et d'aborder en détail le problème de l'héritabilité manquante. Qu'avons-nous appris dans cette plongée dans le cadre interactionniste de la génétique quantitative ?

Toute maladie est potentiellement concernée par la question de l'héritabilité et par la question des variations génétiques : ce point peut sembler trivial, il constitue pourtant une confirmation du processus de généticisation que nous avons décrit dans nos deux premiers chapitres. Il n'y a pas de sens à distinguer entre des maladies génétiques et des maladies non génétiques, sauf à vouloir s'enfermer à nouveau dans le carcan du problème de la sélection causale. Il n'y a pas de sens à distinguer des maladies génétiques parce que l'environnement est un facteur causal dans toutes les maladies (même les maladies dites monogéniques ou mendéliennes ou simples) et il n'y a pas de sens à distinguer des maladies non génétiques parce que les gènes sont un facteur causal dans toutes les maladies (même les maladies communes ou maladies complexes ou maladies polygéniques et même les maladies psychiatriques). La question de l'architecture allélique des maladies souligne d'ailleurs à quel point une

distinction absolue entre maladies mendéliennes et maladies polygéniques est dépourvue de sens : on ne peut pas distinguer entre ces deux types de maladies comme deux types d'architectures avec des variants rares et mendéliens d'un côté et des variants communs à faible effet de l'autre.

Gènes et environnement ne suffisent pas à décrire la matrice causale des maladies : le problème de l'héritabilité manquante souligne bien à quel point la distinction entre gènes et environnement est en partie artificielle et ne suffit pas à décrire la matrice causale des maladies. Parler de « gènes » est une construction pour inclure toutes les sources de variation génétique : les polymorphismes nucléotidiques simples, les duplications ou les délétions (les CNV), les interactions épistatiques, etc. Parler d' « environnement » n'est pas satisfaisant non plus tant le terme est inclusif. Et que faire de l'épigénétique, troisième source de la variation phénotypique selon les biologistes du développement ?

La partition additive entre variance génétique et variance environnementale a des limites : au début de ce chapitre, nous avons rappelé que l'héritabilité est un concept fondé sur une double hypothèse d'additivité (additivité des gènes et de l'environnement, additivité des variations génétiques) mais ces deux hypothèses d'additivité sont fortement remises en cause aujourd'hui, que ce soit au niveau des populations ou au niveau de l'individu. Elles vont de pair avec l'idée qu'il doit être possible de décomposer la variance phénotypique d'une population entre deux sources de variation indépendante l'une de l'autre. Mais ce que nous apprend le concept biologique d'interaction gène x environnement, c'est qu'une telle partition ne fait sens ni du point de vue de l'individu, ni du point de vue des populations puisque cette partition masque le fait que des populations génétiques différentes réagissent de façon différente à une même palette d'environnement. Qu'on soit ou pas d'accord sur l'importance biologique des interactions gène x environnement et de l'épigénétique, il faut reconnaître que la partition additive de la variance phénotypique en variance environnementale et variance génétique est une hypothèse simplificatrice et dont la plausibilité biologique est pour le moins douteuse.

Pour donner une explication du rôle causal des gènes, on ne peut pas se contenter d'une mesure quantitative : quels que soient ses défauts, on peut accorder à l'héritabilité un statut de modèle – certes imparfait et simplificateur – mais qui a permis d'approximer les contributions relatives des gènes et de l'environnement dans la variation phénotypique d'un trait au niveau d'une population. Une telle approximation quantitative est utile, elle a permis d'identifier des variants génétiques impliqués dans les maladies complexes. Mais il est frappant de remarquer à quel point cette partition de la variance et cette approche quantitative se rapprochent de la formulation du problème de la sélection causale tel qu'il a été appliqué à la question des maladies génétiques. Encore une fois, nous ne remettons pas en cause la valeur heuristique d'une formulation du problème de la maladie génétique en ces termes. En revanche, ce dont nous doutons c'est qu'un interactionnisme purement quantitatif, qui considère les gènes et l'environnement comme les deux seuls facteurs de la matrice causale des maladies, et qui considère séparément l'action de chacun de ses facteurs puisse servir de cadre à une explication de la maladie. Ainsi connaître l'estimation de l'héritabilité d'une maladie dans une population donnée ne nous donne *aucune* explication de la maladie. L'estimation de l'héritabilité ne donne bien sûr pas d'explication de la maladie au niveau de l'individu puisque (1) l'héritabilité est un paramètre intrinsèquement dépendant d'une population donnée, (2) l'héritabilité ne nous permet pas de connaître les mécanismes par lesquels un génotype particulier se combine à un environnement particulier pour donner un phénotype particulier. Mais l'estimation de l'héritabilité ne nous donne pas davantage une explication de la maladie au niveau d'une population. En effet, (1) l'héritabilité est un paramètre intrinsèquement dépendant d'une population donnée, de la composition génotypique d'une population, de l'environnement spécifique de la population évaluée – c'est bien une « analyse locale » comme le dit Lewontin qui ne permet pas d'extrapolation hors de la population d'intérêt. Mais surtout (2) l'héritabilité ne fait qu'évaluer quantitativement trois paramètres (variance génétique, variance environnementale, variance phénotypique) qui sont intrinsèquement dépendants de la population. L'héritabilité ne peut ainsi rien nous dire de la façon dont la variance génétique et la variance environnementale se combinent – une telle explication n'aurait littéralement pas de sens au niveau d'une population. En effet, la variance génétique d'une

population recouvre en fait des effets génétiques individuels différents (et éventuellement opposés) pour des groupes génétiques différents de la population et la variance environnementale recouvre en fait des effets environnementaux individuels différents (et éventuellement opposés) pour des groupes de la population. Enfin, l'estimation de l'héritabilité ne prend pas en compte le fait que des populations génétiquement différentes réagissent de façon différente à une même palette d'environnement – ce qu'incarnent les normes de réaction et le concept d'interaction gène x environnement. Pour toutes ces raisons, le cadre interactionniste de la génétique quantitative, s'il permet d'approximer l'influence de la variance génétique et de la variance environnementale sur la variance phénotypique d'une maladie dans une population ne peut en aucun cas servir à l'élaboration d'une explication interactionniste de la maladie. C'est du reste la position qu'expriment les partisans de la théorie des systèmes du développement :

> « La réponse standard à l'opposition nature / culture est le sermon selon lequel, de nos jours, tout le monde est interactionniste : tous les phénotypes sont les produits combinés des gènes et de l'environnement. Selon cette version de la position conventionnelle « interactionniste », le vrai débat ne devrait plus être de savoir si un trait particulier est dû à la nature ou à la culture, mais plutôt à quel point chacune « influence » le trait. La question n'est plus de savoir si l'intelligence est innée ou acquise, mais plutôt de savoir si l'intelligence est génétique à 50 ou à 70%. La DST rejette cette tentative de répartition de la responsabilité causale dans la formation des organismes en des composantes additives. De telles manœuvres ne résolvent pas le débat nature / culture, ils le poursuivent. C'est une façon typique de transformer la vision traditionnelle du développement lorsque celle-ci est défiée et de revenir infester les débats académiques et sociaux actuels. S'il n'est plus aujourd'hui acceptable de se demander si quelque chose est instinctif, alors on se demande s'il a une part génétique importante. Si cela aussi devient inacceptable, alors on demande s'il n'y a pas une prédisposition génétique. Ce dont nous avons besoin c'est d'un mouvement « pieu dans le cœur », c'est d'une façon de penser le développement qui ne se repose pas sur une

> distinction entre des causes essentielles privilégiées et des causes qui sont seulement des supports ou des interférences ». (Oyama, 1985, p. 27)

Ce mouvement « pieu dans le cœur » que décrit Susan Oyama, c'est celui qui nous permet de dégager l'explication du rôle causal des gènes dans la maladie du problème de la sélection causale et du cadre interactionniste quantitatif de l'héritabilité. Mais il s'agit aussi d'éviter de tomber dans ce « consensus interactionniste » trivial que nous avons évoqué au début de ce chapitre et que Susan Oyama rappelle au début de cette citation. Rappelons que le but de ce chapitre est de parvenir à déterminer le cadre interactionniste dans lequel une explication du rôle causal commun des gènes dans les maladies pourrait être légitime. C'est pourquoi, nous allons à présent nous tourner vers une troisième conception de l'interactionnisme, l'interactionnisme « constructiviste », « co-constructionniste » des partisans de la théorie des systèmes en développement.

3.4. L'interactionnisme co-constructionniste

Avant de décrire la forme de l'interactionnisme soutenu par la théorie des systèmes en développement (DST), il est nécessaire d'apporter deux précisions – une précision de terminologie et une précision de méthode. Tout d'abord, d'un point de vue terminologique, nous empruntons le terme d' « interactionnisme co-constructionniste » à Thomas Pradeu. Thomas Pradeu a préféré ce terme à celui de « interactionnisme constructif » choisi par Susan Oyama (Oyama, 1985) ou à celui de « dialectique » des interactions organisme-environnement qu'utilise Lewontin (Levins et Lewontin, 1985), dans le but d'insister sur la relation d'interpénétration entre l'organisme et l'environnement qui est au cœur de la pensée de Lewontin. Par ailleurs, d'un point de vue méthodologique, il est utile de préciser que, contrairement à l'interactionnisme de la génétique quantitative, cet interactionnisme co-constructionniste n'a jamais été appliqué à notre connaissance directement à l'explication des maladies. Cela ne signifie pas qu'il n'y soit pas applicable, simplement qu'il n'a pas été pensé pour ce cadre spécifique. C'est aussi ce qui justifie notre choix

d'analyser plus rapidement cette thèse co-constructionniste[25] que nous ne l'avons fait pour l'interactionnisme de la génétique quantitative pour lequel la maladie est l'un des objets directs d'investigation et d'application.

3.4.1. Contexte et objet de l'interactionnisme co-constructionniste

3.4.1.1. Internalisme développemental et externalisme évolutif

Pour éclairer le débat sur la construction des organismes, nous voudrions reprendre la distinction identifiée par Godfrey-Smith entre deux thèses extrêmes : l'internalisme développemental et l'externalisme évolutif (Godfrey-Smith, 1996). L'internalisme développemental est en quelque sorte l'héritier du préformationnisme, thèse sur le développement des organismes au dix-huitième siècle, selon laquelle toutes les potentialités de l'homme adulte sont déjà préfigurées dans la cellule œuf humaine et dont l'image la plus populaire est celle de l'homoncule. L'internalisme développemental a trouvé un nouveau souffle avec la métaphore du programme génétique, aujourd'hui profondément critiquée (Griffiths, 2001) : l'ADN (et éventuellement la machinerie cellulaire responsable des processus de transcription et de traduction) contiendrait ainsi toute « l'information », toutes les « instructions » qui sont nécessaires pour « coder » les protéines qui vont participer au développement d'un organisme. Dans une telle conception, l'environnement joue bien un rôle dans le développement des caractéristiques individuelles d'un organisme – mais un rôle mineur, un rôle de soutien et de support, un rôle passif :

> « L'internalisme affirme donc que l'organisme est le produit endogène de potentialités qu'il contient en lui dès un stade précoce de son existence. Bien entendu, les partisans de l'internalisme ne nient pas que l'environnement puisse avoir une influence sur la construction de l'organisme à travers le temps, mais l'environnement est seulement vu comme la source de matières premières (par exemple la nourriture, le dioxygène pour la respiration, etc.), ou de déclencheurs (par exemple, certaines plantes désertiques conservent

[25]Pour une description plus détaillée de la thèse de l'interactionnisme co-constructionniste, on se reportera aux travaux de Thomas Pradeu (voir, en particulier, Pradeu, 2009, 2012).

> leurs graines inactives jusqu'à ce qu'une forte chute de pluie se produise, déclenchant alors le développement de la graine), ou encore de perturbations (certaines conditions extérieures, la température par exemple, feraient que le programme interne pourrait plus ou moins bien se réaliser), vis-à-vis d'un développement endogène et autonome. » (Pradeu, 2007, p. 378-379)

À l'opposé de l'internalisme développemental, on trouve l'externalisme évolutif, qui repose sur l'adaptationnisme que Gould et Lewontin ont critiqué dans leur célèbre article de 1979 (Gould et Lewontin, 1979). La façon dont les êtres vivants semblent « convenir » et « correspondre » à leur environnement, provoque traditionnellement l'émerveillement et l'invocation du pouvoir créateur divin pour expliquer cette harmonie flagrante. La métaphore de l'adaptation, couplée à celle de la sélection naturelle et telle qu'elle a été introduite par Darwin, permet d'ailleurs d'expliquer par le même mécanisme aussi bien cette concordance entre les traits phénotypiques des organismes et les caractéristiques de leur environnement que la diversité des organismes et des espèces. L'environnement a certaines caractéristiques ; les organismes dont l'équipement anatomique, physiologique et métabolique s'accorde le mieux à cet environnement ont plus de chances de se reproduire et si ces avantages sont héritables, ils les transmettent à leurs descendants. De ce processus résulte l'augmentation d'une population « mieux adaptée » à son environnement. Si cette même population est répartie dans des conditions spatio-temporelles et physiques différentes, elle acquiert alors différents traits correspondants à ces environnements différents et cette population se différencie en plusieurs espèces ou en sous-espèces. Des adaptations différentes des organismes à des environnements distincts permettent de comprendre l'origine des espèces. Ainsi, l'optimisation des traits phénotypiques des individus comme la diversité du monde vivant est intimement liée à l'idée que les organismes, s'ils veulent survivre aux changements de l'environnement, doivent s'adapter ou mourir car la sélection naturelle ne retiendra que les organismes présentant les meilleures adaptations. La métaphore de l'adaptation peut donc avoir tendance à imposer une vision passive de l'organisme vis-à-vis d'un environnement hyperactif et surpuissant. Un problème est posé par l'environnement à l'organisme qui doit trouver une forme adéquate que

l'environnement sanctionnera ; c'est la métaphore de la clef qui doit s'adapter à la serrure : *« Les organismes s'adaptent à l'environnement parce le monde extérieur a acquis ses propriétés indépendamment de l'organisme qui doit nécessairement s'adapter ou mourir »* (Lewontin, 2000, p. 43).

L'internalisme développemental ou l'externalisme évolutionnaire concourent tous les deux à imposer une séparation radicale entre « l'intérieur » et « l'extérieur » et à mettre l'accent sur l'un ou sur l'autre de ces déterminants pour expliquer la construction de l'organisme. C'est précisément cette dichotomie radicale entre intérieur et extérieur et la prédominance soit des gènes, soit de l'environnement, dans le développement et l'évolution des organismes que critiquent Lewontin et les partisans de la théorie des systèmes en développement, qui affirment au contraire l'interpénétration dynamique et perpétuelle entre un organisme et son environnement. La thèse co-constructionniste est donc double comme le fait remarquer Thomas Pradeu : elle implique d'une part que l'organisme construit son environnement et d'autre part que l'environnement construit l'organisme.

3.4.1.2. L'organisme construit son environnement et l'environnement construit l'organisme

Lewontin résume la première partie de la thèse co-constructionniste (l'organisme construit son environnement) en cinq thèses différentes (Lewontin, 2000) que nous nous contenterons d'énumérer étant donné les limites de ce travail : *(1) Les organismes déterminent quels sont les éléments du monde extérieur qui constituent leur environnement spécifique ainsi que les relations qui sont pertinentes pour eux parmi celles que ces éléments entretiennent. (2) Les organismes ne déterminent pas seulement quels aspects du monde extérieur ont une importance pour eux au vu des particularités de leurs formes, de leurs activités et de leurs métabolismes, mais ils construisent activement, au sens littéral du terme, un monde autour d'eux. (3) Les organismes ne se contentent pas de déterminer ce qui est pertinent et de créer un réseau de conditions physiques parmi les éléments du monde extérieur, mais ils sont aussi en constante altération de leur environnement. (4) Les organismes modulent les proportions statistiques des paramètres de leur environnement. (5) Les organismes*

déterminent par leur biologie la nature physique réelle des signaux en provenance de l'extérieur. Lewontin interprète ainsi les phénomènes de transduction du signal : un signal du monde extérieur physique n'a pas le même sens pour tous les organismes et sa traduction sera différente aussi bien d'un point de vue physique que d'un point de vue comportemental en fonction du métabolisme qui le perçoit.

D'autre part, et c'est la deuxième partie de la thèse, l'environnement construit son organisme. En effet, du point de vue du développement, l'organisme ne se développe jamais uniquement en fonction de ses potentialités internes, il est influencé constamment par l'environnement qu'il construit. Du point de vue de l'évolution, l'environnement agit sur l'organisme par le biais des pressions de sélection. Comme le fait remarquer Thomas Pradeu, il pourrait sembler que l'interactionnisme co-constructionniste ne présente guère d'originalité vis-à-vis de l'externalisme évolutif de ce point de vue. Pour comprendre l'originalité de la deuxième thèse (l'environnement construit l'organisme), il faut admettre que l'environnement n'est pas un paramètre fixe, mais au contraire une variable toujours changeante et dont les changements dépendent justement de l'organisme qui le construit :

> « La réponse du partisan de l'interactionnisme co-constructionniste sur ce point est en apparence semblable à celle que fait l'externaliste : l'environnement sélectionne certaines variations exprimées par les organismes. Cependant, la thèse devient originale dès lors qu'on articule le deuxième aspect (c'est-à-dire la proposition selon laquelle l'environnement sélectionne certaines variations exprimées par les organismes) au premier (c'est-à-dire la proposition selon laquelle l'organisme construit son environnement). Le raisonnement consiste à dire que les gènes influent sur les traits des organismes, qui modifient l'environnement, qui à son tour influe sur la répartition des gènes dans la population concernée (et également chez d'autres espèces). L'interactionnisme co-constructionniste se donne donc ici pour tâche de décrire une dynamique écologique complexe, faite d'actions et de rétro-actions. Autrement dit, c'est la causalité enchaînée des deux phénomènes qui aboutit à une interprétation renouvelée du processus de sélection naturelle. » (Pradeu, 2007, 395)

Dès lors, l'évolution ne s'exerce pas seulement sur des organismes, mais plutôt sur ce que Lewontin appelle des « organismes-environnements ». Si cette conception de l'organisme-environnement comme unité sur laquelle s'exerce la sélection peut paraître séduisante, elle n'est pas sans soulever de nombreuses difficultés : comment définir un organisme-environnement ? Une telle entité peut-elle être une unité de sélection ? De façon plus générale, si la portée développementale de la DST est bien reconnue et établie, sa portée évolutive est aujourd'hui très débattue (Pradeu, 2010). Dans la mesure où la portée évolutive de la thèse co-constructionniste et plus généralement de la théorie des systèmes en développement n'est pas au cœur de notre discussion, nous ne développerons pas davantage cette deuxième partie de la thèse de Lewontin et n'en retiendrons que l'idée d'une dynamique entre organisme et environnement et d'un refus de penser la dichotomie intérieur / extérieur comme une dichotomie fixée et stable.

Nous avons esquissé le contexte général dans lequel Lewontin et les partisans de la théorie des systèmes en développement ont défendu la thèse co-constructionniste. Nous allons à présent nous intéresser à une thèse qui soutient et complète l'interactionnisme co-constructionniste : la thèse de la parité causale.

3.4.2. La thèse de la parité causale

Une première caractérisation grossière de la thèse de la parité causale (que Susan Oyama appelle aussi parfois « démocratie causale » (Oyama, 2000)) consiste à dire que les gènes ne sont ni un facteur causal unique ni un facteur causal privilégié du développement et de l'évolution des organismes. Même en adoptant cette formulation relativement peu controversée, il est déjà possible de percevoir que la thèse de la parité causale est multidimensionnelle et que c'est cette multidimensionnalité qui explique en partie les difficultés d'interprétation qu'elle pose et la vaste littérature dont elle fait l'objet – à quoi il faut ajouter que la théorie des systèmes en développement rassemble différents biologistes et/ou philosophes (Griffiths et Gray, 1994, 2005 ; Oyama, 1985, 2000 ; Oyama, Griffiths et Gray, 2001) qui n'acceptent pas toujours les différentes dimensions de la thèse de la parité causale ou qui en donnent des formulations distinctes et enfin, que les opposants ou ceux qui ont

fait preuve d'un certain scepticisme vis-à-vis de la thèse de la parité causale ont encore obscurci un peu plus le débat (Kitcher, 2001 ; Schaffner, 1998). C'est ainsi qu'Ulrich Stegmann distingue pas moins de sept versions différentes de la thèse de la parité causale (Stegmann, 2012).

Cependant, étant donné que nous ne nous intéressons pas à la question du rôle causal des gènes dans les traits en général mais à la question du rôle causal des gènes dans les maladies en particulier, nous estimons qu'il n'est pas nécessaire de détailler ces sept versions et que pour l'usage que nous comptons en faire, trois distinctions suffisent à exposer la thèse de la parité causale. Ces trois distinctions sont : la question du « pied d'égalité des causes » et de la « non-séparabilité », le sens fort ou le sens faible de la théorie de la parité causale et la portée développementale et la portée évolutionniste de la thèse de la parité causale.

De la même manière que Richard Lewontin refuse une dichotomie forte entre organisme et environnement, dans leur article de 1994, Russell Gray et Paul Griffiths refusent une séparation dichotomique entre le rôle causal joué par les facteurs génétiques et celui joué par les facteurs non génétiques dans le développement :

> « La théorie des systèmes en développement rejette l'approche dichotomique du développement : les gènes sont juste une ressource disponible pour le processus du développement. Il y a une symétrie fondamentale entre le rôle des gènes, et celui du cytoplasme maternel ou de l'exposition dans l'enfance au langage. L'ensemble des ressources développementales représente un système complexe qui est répliqué dans le développement. Il y a beaucoup à dire sur les différents rôles des ressources particulières. Mais il n'y a rien qui divise les ressources en deux catégories fondamentales. Le rôle des gènes n'est pas plus unique que le rôle de nombreux facteurs ». (Griffiths et Gray, 1994, p. 277)

La première chose à remarquer ici, c'est que Griffiths et Gray désignent sous le terme de facteurs non génétiques aussi bien le « cytoplasme maternel » qui contient les mitochondries et qui est hérité en même temps que l'ADN maternel lors de la

formation du zygote, que « l'exposition au langage dans l'enfance » qui est un paramètre environnemental extérieur. Griffiths et Gray rompent donc ici une double dichotomie : ils font tomber aussi bien la distinction entre les gènes et l'environnement (compris au sens d'environnement extérieur à l'organisme que la distinction entre les gènes) que la distinction entre les gènes et le reste des ressources internes au développement et qui peuvent donc comprendre l'ARN, les enzymes, les modifications épigénétiques, le cytoplasme, etc. Comme le fait remarquer Philippe Huneman (Huneman, 2013), ce refus des dichotomies entre les rôles des différentes ressources développementales va de pair avec leur inséparabilité : les modifications épigénétiques impliquent une méthylation de l'ADN et il n'y a pas de sens à parler de méthyle indépendamment de l'ADN, l'ADN ne peut se répliquer et passer par les processus de transcription et de traduction sans la totalité de la machinerie cellulaire, etc. Les facteurs environnementaux externes eux-mêmes sont difficilement séparables de l'organisme et c'est en ce sens que l'on peut parler de « systèmes en développement » ou de « système organisme-environnement » comme le fait Lewontin.

Un autre point qu'il est essentiel de souligner dans le discours de Griffiths et Gray c'est que l'ADN n'est pas la seule molécule à se répliquer : c'est « l'ensemble des ressources développementales » qui se répliquent. En effet, les gènes ne sont pas les seules molécules héritées qui varient : il y a d'autres formes de variation héritable telles que l'ADN mitochondrial, certaines modifications épigénétiques – certains vont même jusqu'à parler d'hérédité culturelle pour certains phénomènes sociaux. C'est pourquoi la thèse de la parité causale a non seulement une portée développementale, mais aussi une portée évolutive. Non seulement, elle affirme qu'il y a une symétrie dans le rôle causal que jouent les gènes et les facteurs non génétiques dans le développement, mais elle affirme la même symétrie dans l'évolution.

Mais que veulent dire Griffiths et Gray en affirmant la symétrie du rôle causal de toutes ces ressources développementales avec l'ADN dans le développement et l'évolution ? Il y a au moins deux façons d'interpréter cette affirmation, ce que Marcel Weber a distingué sous les termes de « version forte » et « version faible » de la thèse de la parité causale :

> « Dans la DST, la thèse de la parité causale ne signifie sûrement pas que le rôle de l'ADN ne diffère en aucune façon de celui des autres composants cellulaires. Ce que la DST remet en question, c'est qu'il y ait une différence catégorique dans le rôle de l'ADN – quelque chose comme un contenu informationnel ou un rôle de programmation, de direction, d'orchestration.
>
> Pour répondre à cet argument, il semble nécessaire de distinguer entre une version forte et une version faible de la thèse de la parité causale. La version forte dit que l'ADN et les gènes ne jouent aucun rôle causal qui les distingue des autres composants des systèmes en développement. En d'autres termes, il n'y aucune différence de rôle causal, selon la version forte. La version faible affirme seulement qu'il n'y a pas de différences catégoriques, c'est-à-dire, que même si le rôle causal de l'ADN et des gènes est différent, ils n'appartiennent pas à une catégorie séparée de causes développementale. » (Weber, 2005, p. 260)

La version forte signifie qu'il n'y aurait aucune différence dans le rôle de l'ADN et des autres ressources développementales, alors que la version faible implique que les différences de rôle causal susceptibles d'exister ne justifient pas que l'on considère que l'ADN et les autres ressources fondamentales sont deux causes de catégorie ou de nature différente – l'ADN étant la cause du développement et de l'évolution des organismes, alors que les autres ressources développementales seraient reléguées au rang de condition. La version forte de la thèse de la parité causale a été attaquée à de nombreuses reprises au point que Griffiths et Knight ont précisé leur propos – réfutant la version forte de la thèse de la parité causale comme une « parodie » de leur position :

> « La parité est l'idée que les gènes et les autres causes matérielles sont sur un pied d'égalité. La parodie du développementalisme qui sert d'épouvantail dit que toutes les causes du développement sont d'égale importance. La vraie position des développementalistes est que les différences empiriques entre le rôle de l'ADN et celui des gradients cytoplasmiques ou des évènements d'empreinte de l'hôte ne justifient pas les distinctions métaphysiques que l'on

> construit à l'heure actuelle sur elles. Les séquences d'acides nucléiques et les membranes phospholipidiques ont toutes les deux des rôles essentiels dans la chimie de la vie et il semble qu'aucun substitut réaliste n'existe pour les unes ou les autres. Mais les faits du développement ne justifient pas qu'on assigne à l'ADN le rôle de l'information et du contrôle tandis que les modèles hérités des membranes sont considérées comme le « support matériel » pour la lecture de l'ADN. Les faits de l'évolution ne justifient pas un rôle spécial pour les séquences d'ADN comme « réplicateurs » tandis que les membranes sont regroupées avec tout ce qui, des groupes méthyles à la culture humaine, est considéré comme « des interacteurs ». (Griffiths et Knight, 1998, p. 254)

Les deux dimensions (développementale et évolutive) de la thèse de la parité causale apparaissent bien ici. Il est capital de bien comprendre les distinctions faites ici par Griffiths et Knight. Le vocabulaire des réplicateurs a d'abord été introduit par Dawkins (Dawkins, 1976), par opposition au terme de « véhicule ». Pour clarifier le débat sur les unités de sélection naturelle, David Hull a conservé le terme de « réplicateur », mais a préféré au terme de « véhicule » celui d' « interacteur » (Hull, 1980). Pour David Hull, les philosophes confondent sous un même terme « unité de sélection » deux rôles importants : le rôle de réplicateur – assuré par une entité transmettant sa structure de manière intacte au travers de réplications successives et le rôle d'interacteur – assuré par une entité qui interagit en tant qu'ensemble cohérent avec son environnement, de telle sorte que cette interaction induit une réplication différenciée des divers éléments composant cette entité au sein de la population dans son ensemble. Richard Dawkins a ensuite repris le vocabulaire des réplicateurs et des interacteurs dans son ouvrage de 1982 (Dawkins, 1982). Si Hull et Dawkins s'accordent tous les deux sur le rôle des gènes comme réplicateurs, ils sont en désaccord sur les entités qui supportent le rôle des interacteurs (Godfrey-Smith, 2000). Nous ne détaillerons pas d'avantage cette discussion sur la définition et le rôle des réplicateurs et des interacteurs – il suffit de noter que cette distinction a été utilisée pour attribuer des rôles causaux d'importance différente à des entités comme les gènes, les mitochondries ou les variations épigénétiques dans la théorie de l'évolution. La thèse de la parité causale, telle qu'elle est décrite par Griffiths et Knight ne nie pas

que l'ADN puisse avoir un rôle causal différent des autres facteurs causaux du développement ou de l'évolution. En revanche, la thèse de la parité causale affirme que ces rôles causaux sont différents en fonction de la question que l'on pose (selon qu'on s'intéresse à l'évolution ou au développement) et que la différence de ces rôles causaux ne permet pas d'affirmer la prédominance causale des gènes, que ce soit dans le développement ou dans l'évolution. C'est cela qui est signifié par le terme « sur un pied d'égalité » et non pas que les gènes et les autres facteurs causaux sont des causes « d'égale importance ». Cette confusion a souvent été faite cependant, ce qui explique qu'en 2005, Griffiths et Gray reviennent sur ce point dans un article dont l'objectif est de dénoncer trois malentendus sur la théorie des systèmes en développement :

> « Le troisième malentendu identifie la théorie des systèmes en développement avec la thèse selon laquelle « les gènes ne sont pas importants » ou « les gènes sont moins importants que ce que les gens pensent ». La conséquence de ceci, c'est que la DST est vue comme simplement une autre critique de la « génétique du sac de haricots », cette critique ayant perdu sa pertinence à partir du moment où la génétique du développement moderne s'est sophistiquée (Gilbert, 2003). Ceci, cependant, révèle une incompréhension fondamentale de la « thèse de la parité », selon laquelle les rôles joués par les nombreux facteurs causaux qui affectent le développement ne correspondent pas à une distinction nette entre deux catégories, un rôle exclusivement joué par les éléments de l'ADN et un autre exclusivement joué par les éléments qui ne sont pas de l'ADN. » (Griffiths et Gray, 2005, p. 420)

3.4.3. La thèse de la parité causale est-elle applicable ?

Si les partisans de la théorie des systèmes en développement ont dû revenir plusieurs fois sur la signification exacte de la thèse co-constructionniste et de la thèse de la parité causale, c'est probablement parce que ces deux thèses, aussi séduisantes soient-elles, ont été souvent considérées comme peu applicables et trop programmatiques. A force de refuser les dichotomies et les distinctions, la théorie des systèmes en développement risquerait de se diluer dans un « holisme ingérable »

(Sterelny *et al.*, 1996). C'est un constat relativement similaire que fait Philip Kitcher lorsqu'il discute de la portée de la théorie des systèmes en développement contre le déterminisme génétique (Kitcher, 2001, p. 408). Alors même que Philip Kitcher reconnaît l'intérêt théorique de la DST et de la thèse co-constructionniste, il considère qu'elle se limite soit à une critique d'une biologie moléculaire qui a déjà évolué (aucun biologiste ne se reconnaîtrait dans l'affirmation que les gènes sont les conditions nécessaires et suffisantes au développement d'un trait), soit à un programme encore trop peu défini pour avoir des applications pratiques dans notre compréhension du développement. Des critiques similaires, peut-être encore plus insistantes, ont été adressées à la portée évolutive de la DST. La DST insiste sur le fait que l'ADN n'est pas le seul système d'hérédité : d'autres systèmes existent qui doivent être pris en compte. Néanmoins, Griffiths et Gray sont assez réticents au concept « d'hérédité multiple » (Griffiths et Gray, 2001, p. 198). Cette opposition de Griffiths et Gray à l'utilisation des termes « hérédité multiple » ou « réplicateur étendu » est liée à la thèse de la non-séparabilité qui va de pair avec celle de la parité causale. C'est parce que les systèmes d'hérédité ont tous le même statut de cause du développement et que considérer les systèmes d'hérédité de façon séparée ne peut avoir qu'une valeur purement heuristique et ne durer que le temps de la recherche que Griffiths et Gray préfèrent le terme de « systèmes en développement ». Cependant, le maintien de cette distinction aurait selon Francesca Merlin des conséquences fâcheuses. A force de refuser de considérer les systèmes d'hérédité comme séparés, on ne peut plus faire sens de la thèse de la parité causale qui, tout en reconnaissant le statut de cause à toutes les ressources développementales, est supposée admettre la reconnaissance de rôles causaux différents :

> « L'idée holiste de Griffiths et Gray selon laquelle les organismes héritent la matrice environnementale entière est certainement plus riche et plus complète que les descriptions alternatives de l'hérédité limitées à un petit ensemble de facteurs héréditaires qui interagissent de façon additive. Cependant, je défends que, si nous comprenons l'hérédité comme un système global composé de multiples sous-systèmes quasi indépendants d'hérédités qui interagissent de façon non additive (ce n'est pas la façon dont Griffiths et

> Gray représentent le système de l'hérédité), il est plus utile et biologiquement réaliste de maintenir leurs distinctions relatives afin de comprendre leurs interactions et leurs influences réciproques. L'approche holiste de Griffiths et Grays est finalement moins informative que la représentation de l'hérédité que je défends, simplement parce qu'elle nous dit que les organismes héritent d'un système très complexe d'interactions non-linéaires mais qu'elle ne nous dit rien de plus à propos de la dynamique qui permet à un tel système d'être hérité (de ré-advenir ou d'être reconstruit à chaque génération, selon les termes de Griffiths et Gray). » (Merlin, 2010, p. 210)

Il nous semble cependant que la critique de Francesca Merlin est peut-être un peu radicale et qu'on peut interpréter l'adhésion de Griffiths et Gray à la thèse de la non-séparabilité comme une attitude pragmatique : comme Susan Oyama l'a bien montré, il est possible de faires différentes distinctions opératoires entre les différentes ressources développementales en fonction des programmes de recherche ; il n'y a en revanche pas de distinction qui soit valable quel que soit le contexte explicatif considéré (Oyama, 2000). Dès lors, que retenir de l'interactionnisme co-constructionniste et de la thèse de la parité causale dans notre perspective qui cherche à déterminer le rôle causal commun des gènes dans les maladies ?

3.4.4. Que nous apprennent l'interactionnisme co-constructionniste et la thèse de la parité causale ?

Nous l'avons déjà dit, l'interactionnisme co-constructionniste et la thèse de la parité causale ont pour objet principal la construction des caractéristiques individuelles de l'organisme, dans son développement et dans son évolution. Ces deux thèses ne sont donc pas directement appliquées à la question des maladies – contrairement à l'interactionnisme de la génétique quantitative. Ceci dit, rien n'interdit *en théorie* que l'interactionnisme co-constructionniste et la thèse de la parité causale puissent s'appliquer aux maladies, soit que l'on considère que la maladie est un trait phénotypique comme un autre exprimé par un organisme à un moment de son développement, soit que l'on considère que la maladie est un événement qui perturbe le système en développement à un moment donné de son cycle de vie. Au contraire,

l'interactionnisme co-constructionniste et plus encore la thèse de la parité causale affirment qu'il n'y a pas de raison de donner une prédominance causale aux gènes dans le développement des traits d'un organisme et qu'il n'y a pas de raison de distinguer comme des causes de nature différente les gènes des facteurs non génétiques. En ce sens, la théorie de la parité causale s'oppose donc frontalement au problème de la sélection causale qui nous oblige à trouver des justifications au rôle causal prédominant des gènes et dont nous avons établi plus tôt qu'il menait à une impasse. Mieux encore, au vu de ce que l'on a reproché à l'interactionnisme de la génétique quantitative (description binaire et insuffisante de la matrice causale des maladies, dichotomie artificielle entre gènes et environnement, refus de prendre en compte les différents systèmes d'hérédité, difficulté à rendre compte des interactions entre les différentes causes des maladies), il semblerait même que la théorie co-constructionniste et la thèse de la parité causale conviennent parfaitement au cadre interactionniste qualitatif qui était préalablement nécessaire à une discussion du rôle causal commun des gènes dans toutes les maladies.

Cependant, si l'on veut éviter que ce que nous gagnons en perception de la complexité de la causalité des maladies soit perdu dans ce que Jensen appelait péjorativement « l'interactionnisme » (c'est à dire une causalité trop complexe pour être déchiffrée), il paraît essentiel *en pratique* de distinguer les systèmes d'hérédité ou les différentes ressources développementales entre elles, au cas par cas, dès qu'on analyse une explication interactionniste co-constructionniste des maladies.

3.5. Conclusion du chapitre : l'hypothèse d'une théorie génétique de la maladie

À la fin de la première partie de ce travail, nous avons choisi d'abandonner le projet de définir le concept de maladie génétique pour lui préférer celui de donner un sens ni trivial ni génocentriste au processus sociologique de généticisation. Pour ce faire, il nous fallait définir un cadre interactionniste qui soit adapté à la détermination du rôle causal des gènes dans le développement de toutes les maladies. L'objectif de ce chapitre était donc de se frayer un chemin au milieu des multiples concepts d'interactionnisme qui existent en biologie.

Nous avons d'abord exposé ce que nous avons appelé « l'interactionnisme vague », aussi appelé « consensus interactionniste » ou « interactionnisme biologique standard », au motif que si cet interactionnisme est unanimement accepté, sa trivialité même lui fait perdre toute dimension explicative. Nous avons poursuivi par la description du cadre interactionniste de la génétique quantitative, tel qu'il s'est construit autour du concept d'héritabilité. Nous en avons conclu que si l'héritabilité était un concept bien vivant et qui avait permis l'élucidation d'un certain nombre de variants génétiques impliqués dans les maladies, le problème de l'héritabilité manquante soulignait cruellement les insuffisances de ce modèle. Cet interactionnisme, bien qu'il soit applicable à toute maladie, se fonde en effet sur une partition additive de la variance phénotypique d'un trait dans une population et reproduit les impasses auxquelles la formulation du problème de la maladie génétique en termes de problème de la sélection causale nous avait déjà menée : description binaire et insuffisante de la matrice cause des maladies, construction artificielle d'une dichotomie entre gènes et environnement, incapacité à rendre compte des interactions biologiques qui ont lieu entre les différents facteurs causaux du développement des maladies, que ce soit les interactions entre les gènes ou les interactions gène-environnement.

À la recherche d'un interactionnisme qualitatif et non quantitatif, nous nous sommes alors tournée vers l'interactionnisme co-constructionniste et la thèse de la parité causale, tels qu'ils ont été développés par les partisans de la théorie des systèmes en développement. Ce cadre interactionniste présente sans aucun doute de nombreuses qualités. Il permet ainsi de décrire de façon beaucoup plus complète la matrice causale des traits (et donc en ce qui nous concerne des maladies) en incluant à la fois les autres facteurs internes du développement des traits et les facteurs environnementaux. Il remet en cause la dichotomie entre gènes et environnement et s'oppose à l'établissement de la prédominance causale des gènes, en considérant que les gènes comme les autres facteurs non génétiques sont des causes de même nature et que le choix de considérer que l'un ou l'autre sont fixés est purement artificiel. Il admet cependant que les gènes puissent avoir un rôle causal différent en fonction du contexte des autres facteurs non génétiques. C'est donc un cadre interactionniste qui semble répondre aux problèmes soulevés par l'examen de l'interactionnisme de la

génétique quantitative. Il reste pourtant une difficulté de taille : le risque de l'interactionnisme co-constructionniste, c'est de tenir à ce point à une description complexe des systèmes en développement que celle-ci pourrait bien perdre tout contenu informatif et nous ramener au « consensus interactionniste » que nous cherchions à éviter.

L'enjeu de la suite de ce travail est donc clair : il s'agit d'adopter la thèse co-constructionniste et celle de la parité causale comme un cadre général de pensée tout en distinguant au moins d'un point de vue heuristique le rôle causal des gènes du rôle des autres facteurs non génétiques dans le développement des maladies afin de spécifier en quel sens les gènes jouent un rôle causal dans toutes les maladies. Ce rôle causal commun des gènes qui n'implique aucune prédominance causale, c'est ce que nous appelons une théorie génétique des maladies et c'est la façon dont il faut comprendre ce terme que nous allons développer dans le chapitre suivant.

Chapitre 4 : Établir des critères opérationnels pour identifier une théorie génétique

La première partie de ce travail nous a permis de conclure que la généticisation des maladies ne pouvait être réduite à une survivance du génocentrisme. Ce constat nous a amenée à formuler l'hypothèse selon laquelle le processus sociologique de la généticisation pourrait bien refléter un processus épistémique – à savoir l'élaboration progressive d'une explication du rôle causal commun des gènes dans toutes les maladies. Pour tester cette hypothèse, il faut d'abord déterminer les conditions de possibilité de cette explication du rôle causal commun des gènes dans toutes les maladies. Dans le chapitre 3, nous avons établi le cadre interactionniste dans lequel une explication du rôle causal commun des gènes ne serait ni triviale, ni génocentriste, ni un retour déguisé au problème de la sélection causale. Nous avons pu ainsi affiner notre hypothèse – la généticisation pourrait refléter une explication du rôle causal commun des gènes, que nous avons appelée « une théorie génétique » de la maladie – et identifier une première condition de possibilité de cette hypothèse : il ne peut y avoir de théorie génétique de la maladie que dans le cadre d'une explication interactionniste co-constructionniste des maladies.

Une démarche minimalement (et transitoirement) normative

Il s'agit maintenant dans ce chapitre 4 de clarifier ce que nous entendons précisément par « théorie génétique ». Notre position est ici explicitement normative – non pas au sens du débat entre normativisme et naturalisme en médecine, mais au sens le plus classique du terme en philosophie des sciences. Il s'agit d'examiner les conditions de possibilité d'une théorie génétique de la maladie, d'imaginer ce à quoi pourrait ressembler une théorie génétique, si elle existait : quelles prémisses et quelles conditions sont nécessaires pour pouvoir parler de théorie génétique de la maladie ? Qu'est-ce que ce type de théorie devrait expliquer ? A quelle forme d'explication correspondrait une théorie génétique de la maladie ?

Précisons tout d'abord que cette position explicitement normative sera aussi minimalement normative : il ne s'agit pas de proposer une version trop exigeante du

concept de théorie génétique mais de parvenir à des critères opérationnels qui permettent d'identifier ce qui constituerait une théorie génétique *a minima*. Autrement dit, nous ne cherchons pas à parvenir à un modèle unique de théorie génétique mais bien à un spectre des formes (de la théorie génétique *a minima* à la théorie génétique *a maxima* ou idéale) que pourrait prendre une telle théorie.

Par ailleurs, précisons dès à présent que cette position minimalement normative ne constituera qu'un moment de notre réflexion : la troisième partie de cette thèse consistera en effet à adopter une position descriptive-critique, c'est-à-dire à confronter les conditions de possibilité d'une théorie génétique de la maladie à ce qui peut s'apparenter explicitement ou implicitement à des théories génétiques de la maladie dans la littérature biomédicale contemporaine.

Justification méthodologique

Pour établir les conditions de possibilité d'une théorie génétique, nous avons choisi de proposer un long détour par le concept de théorie médicale et de développer notre propre conception de ce que peut et doit être une théorie médicale. Cette démarche peut doublement surprendre. Premièrement, pourquoi faire un détour par le concept de théorie médicale pour penser le concept de théorie génétique ? A cette première question, nous répondons que la nécessité de penser le concept de théorie génétique en relation avec la multitude des explications en médecine est évidente, si l'on veut éviter de retomber dans les affres du génocentrisme. Puisqu'une théorie génétique de la maladie ne saurait être l'affirmation que les gènes sont la cause la plus importante de toutes les maladies, il nous faut préciser en quel sens une théorie génétique est une théorie médicale partielle, qui peut et doit s'intégrer à une explication plus générale de la maladie.

Deuxièmement, certains auteurs comme Kenneth Schaffner, Paul Thompson, Kazem Sadeg-Zadeh et Paul Thagard ont déjà proposé une exploration du concept de théorie médicale : pourquoi ne pas avoir repris une de leurs conceptions ? A cette seconde question, nous objectons (et nous le démontrerons au cours de ces sections) que si chacune de ces conceptions contribue à éclairer le concept de théorie médicale,

aucune ne parvient simultanément à a) identifier la spécificité des théories médicales[26] par rapport aux théories biologiques, b) proposer un concept de théorie médicale qui permette de rendre compte du passage de l'élaboration de la théorie médicale dans la recherche à son application clinique dans la pratique quotidienne du médecin confronté à des situations particulières, c) rendre compte à la fois de la diversité des niveaux d'explication et de la variété des types d'explications qui coexistent en médecine, d) proposer des critères opérationnels qui rendent identifiables ces théories médicales dans la littérature scientifique.

Plan du chapitre et annonce des principales conclusions

Pour parvenir à définir les critères qui permettent d'identifier une théorie génétique, nous procéderons en trois temps.

Dans une première section, il s'agit de justifier pourquoi nous avons choisi de formuler notre hypothèse sur le statut de la généticisation des maladies en termes de « théorie » génétique de la maladie (et non pas, par exemple, en termes de « modèle génétique » de la maladie). Nous précisons en quel sens nous adoptons une conception déflationniste du concept de théorie scientifique. Nous appelons « théorie » une hypothèse scientifique, fondée sur des prémisses empiriques, qui permet d'unifier sous une même explication plusieurs classes de faits, qui étaient expliqués jusqu'ici de façon distincte.

Dans les sections 2, 3 et 4, nous nous proposons de repenser le concept de théorie médicale en passant en revue successivement trois questions : Quel est l'objet des théories médicales ? Comment une théorie médicale explique-t-elle ? Et jusqu'où une théorie médicale doit-elle expliquer ? Ces sections nous permettent de distinguer différents niveaux de théorie médicale (théorie générale de la maladie, théorie régionale d'une classe de maladie, théorie d'une maladie individuelle), différentes cibles explicatives d'une théorie médicale (causes, symptômes, évolution, traitement), différents types d'explication possibles (explication épidémiologique, explication mécaniste, explication évolutionnaire) et des degrés d'exhaustivité explicative

[26] Schaffner et Thompson ne distinguent ainsi pas les théories biologiques des théories médicales et parlent de « théories biomédicales » comme d'un ensemble homogène

différents (théorie *a minima*, théorie *a maxima*, théorie partielle). À partir de ces distinctions qui constituent autant de critères opérationnels, nous considérons qu'une théorie médicale *a minima suppose au minimum un type d'explication (évolutionnaire, épidémiologique, ou physiopathologique) d'au minimum une cible explicative (causes, symptômes, évolution, ou traitement) d'une maladie individuelle en tant que maladie-type, ou d'une classe de maladie ou de la maladie en général.* Pour obtenir le spectre des formes de théories médicales possibles, il suffit de combiner ces critères : certaines théories médicales expliquent une maladie type, une classe de maladie ou la maladie en général ; certaines théories ne font appel qu'à un seul type d'explication tandis que d'autres intègrent plusieurs types d'explication ; certaines théories n'expliquent qu'une cible explicative d'une maladie (ex : l'évolution de la maladie) tandis que d'autres les expliquent toutes. Nous identifions également trois classes de faits qu'il est important d'expliquer en médecine et qui permettent de juger de la valeur intrinsèque d'une théorie médicale et de comparer deux théories médicales de même niveau entre elles. Il s'agit de la variabilité inter-individuelle, de la variabilité intra-individuelle et de la comorbidité.

Enfin, dans la section 5, nous revenons au concept de théorie génétique. En transposant les critères que nous avons établis pour identifier à la fois une théorie médicale *a minima* et le spectre des théories médicales possibles, nous proposerons une typologie des formes possibles de théories génétiques que pourrait refléter la généticisation. Nous distinguerons en particulier deux hypothèses : la généticisation comme *théorie génétique des maladies* et la généticisation comme *théorie génétique de la maladie*.

4.1. Pourquoi faire l'hypothèse d'une « théorie » génétique de la maladie ?

Nous avons formulé notre hypothèse sur le statut épistémique de la généticisation des maladies en termes de « théorie génétique ». Une telle formulation peut surprendre tant le concept de « théorie scientifique » est multiplement connoté dans l'histoire de la philosophie des sciences. Pourquoi avoir formulé notre hypothèse sur la généticisation en termes de « théorie », alors que des termes comme ceux de « modèles » (voir par exemple Suppes, 1961 ; Giere, 2004 ; Hartmann et Frigg, 2006),

de « paradigme » (Kuhn, 1962) ou de « programme de recherche » (Lakatos, 1976, 1980) sont pourtant préférés aujourd'hui par une large partie des philosophes des sciences, soit parce qu'ils sont moins historiquement connotés (c'est le cas du concept de modèle), soit parce qu'ils semblent connotés de façon moins exigeante (c'est-à-dire qu'adopter le terme de « paradigme » ou de « programme de recherche » semble faire appel à des prémisses moins fortes sur la structure de la science) ?

Trois raisons principales ont motivé notre choix : (1) Le terme de « théorie » est régulièrement utilisé dans l'histoire de la médecine, de la « théorie des humeurs » hippocratico-galénique à la « théorie des germes ». (2) Des concepts comme ceux de « modèles » et de « paradigme » nous paraissent poser autant de difficultés définitionnelles que le concept de théorie. (3) Enfin, nous adoptons une conception déflationniste du concept de théorie scientifique, qui n'implique aucunement une filiation de pensée avec le cercle de Vienne, c'est-à-dire avec une conception unifiée de la science dont toutes les branches pourraient être traduites dans une langue universelle et où toutes les théories pourraient être pensées sur le modèle des théories physiques.

4.1.1. Rejet de la conception syntaxique des théories scientifiques

Puisant ses racines dans les idées du cercle de Vienne, la conception syntaxique des théories scientifiques s'est développée à partir de la description des théories physiques pour atteindre son point d'orgue dans les années 1960 (Hempel, 1965 ; Nagel, 1961). Elle repose sur une définition des théories comme un ensemble de lois interconnectées par des relations souvent formalisées en langage mathématique et qui relient des entités théoriques entre elles. Cet ensemble de lois permet de rendre compte des régularités empiriques observables dans la nature. Ce qui permet aux lois d'expliquer les régularités empiriques, aux entités théoriques de décrire les phénomènes observables, c'est l'utilisation de règles de correspondance ou « principes-pont ». Or, si cette conception convient plutôt bien aux théories physiques, elle semble exclure de fait l'existence de théories en biologie, tant il est difficile de pouvoir identifier des lois aussi strictement définies en biologie, où les généralités universelles sont rares et où les régularités empiriques ne se prêtent pas toujours à une formalisation mathématique (Ruse, 1973 ; Smart, 1963). C'est pourquoi les

philosophes de la biologie ont plutôt eu tendance à adopter une conception sémantique des théories scientifiques où les théories sont considérées comme des familles ou des classes de modèles (Suppe, 1972, 1977 ; Suppes, 1967).

Nous ne reprendrons pas ici l'ensemble des arguments qui s'opposent à une conception syntaxique des théories scientifiques en biologie et qui plaident en faveur d'une conception sémantique, d'une part parce qu'il existe déjà d'excellentes synthèses de ces discussions (par exemple, Thompson, 1989) et d'autre part parce que ce point fait largement consensus aujourd'hui en philosophie de la biologie. Ajoutons seulement que si la conception syntaxique a été écartée pour la description des théories biologiques au motif que peu de généralités en biologie sont susceptibles de constituer des lois, on comprend bien qu'*a fortiori*, les quelques auteurs qui se sont intéressés au concept de théorie médicale sont majoritairement des tenants de la conception sémantique, comme le remarque Maël Lemoine (Lemoine, 2013) en évoquant les travaux récents de Kenneth Schaffner (Schaffner, 1993, 1998, 2008), Paul Thompson (Thompson, 1989, 2011a, 2011b) et Kazem Sadegh-Zadeh (Sadegh-Zadeh, 2008, 2011).

Le seul auteur récent qui propose une conception quelque peu différente des théories médicales, une conception dite « cognitive », est Paul Thagard (Thagard, 1998, 2000, 2005, 2008). Encore celui-ci reconnaît-il que la conception cognitive ne s'oppose pas directement à la conception sémantique mais ajoute à une conception des théories scientifiques fondée sur la description de familles de modèles, la dimension psychologique de ces modèles et des agents qui les utilisent pour décrire et expliquer des phénomènes observables. La suite de notre exposé sera donc compatible avec toute conception des théories médicales qui s'est affranchie de la conception syntaxique des théories scientifiques.

4.1.2. Examen des critères communément évoqués pour définir une théorie scientifique

Sans prétendre recenser l'ensemble des définitions qui ont été proposées pour le concept de théorie scientifique, certains critères semblent récurrents comme le concept d'explication causale, les concepts d'universalité ou de nécessité, la notion de pouvoir prédictif, la cohérence interne (de la théorie elle-même) ou externe (avec

l'ensemble des connaissances du champ auquel la théorie se rapporte), les notions de vérifiabilité ou de falsifiabilité. Le concept d'explication causale nous paraît doublement trop restrictif : d'abord parce que le concept de cause en médecine est fortement polysémique (nous reviendrons sur ce point dans le 4.3), mais aussi parce qu'adopter ce concept comme critère principal de définition d'une théorie reviendrait à exclure des théories médicales toute explication de l'évolution d'une maladie ou de ses symptômes (nous détaillerons ce point dans le 4.2.3.). Les critères d'universalité et de nécessité semblent également à proscrire pour les raisons que nous avons évoquées plus haut et qui tiennent à l'objet des sciences biomédicales : même des théories biologiques solides comme la théorie de l'évolution ne peuvent souscrire aux critères d'universalité et de nécessité, tant il est vrai qu'il y a peu de lois en biologie et que les théories biologiques ne portent souvent que sur un sous-ensemble d'êtres vivants ou ne sont nécessaires que si l'on prend en compte des conditions *ceteris paribus*. Enfin des critères comme les notions de cohérence ou de consistance sont sources de confusion : ces critères ne constituent pas des critères de définition qui permettraient de distinguer une théorie d'une hypothèse scientifique (ce qui est l'objet de notre propos) mais des critères d'évaluation de la pertinence ou de la force d'une théorie scientifique (ce à quoi nous nous intéresserons dans un second temps).

4.1.3. Une conception déflationniste des théories scientifiques – critères de Laplane

C'est pourquoi nous nous proposons à la suite de Lucie Laplane (Laplane, 2013, 2014), d'utiliser le vieux principe de *vera causa* pour identifier les théories scientifiques et distinguer dans un premier temps une théorie scientifique au sens strict d'une simple hypothèse. Le principe de *vera causa* a une longue histoire dans les querelles scientifiques (Kavaloski, 1974) et il n'est généralement pas utilisé pour l'identification des théories scientifiques. Mais en utilisant l'interprétation par Darwin du principe de *vera causa* dans le cadre d'une controverse avec ses détracteurs et en reprenant l'analyse que Jean Gayon propose de l'utilisation du principe de *vera causa* chez Darwin (Gayon, 1993, 1997, 2009), Laplane propose trois critères pour identifier une théorie scientifique (Figure 13) :

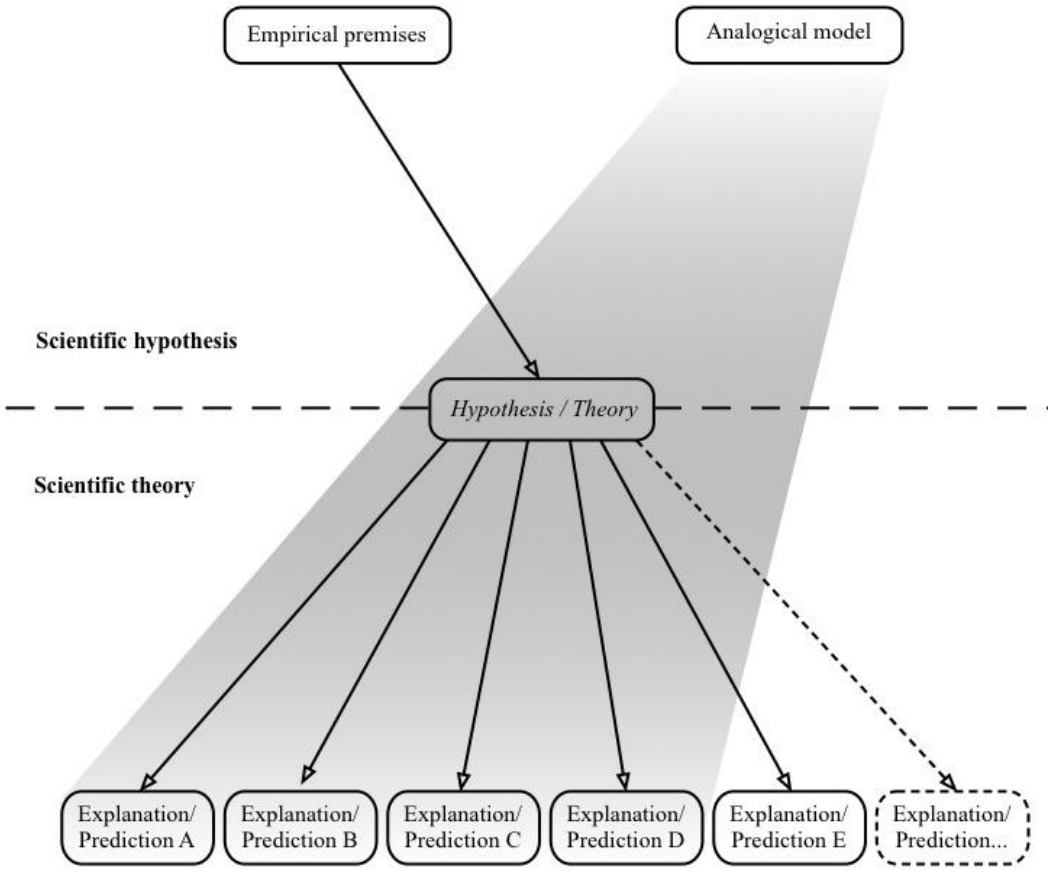


Figure 13 : De l'hypothèse scientifique à la théorie scientifique (Laplane, 2013) : une hypothèse scientifique s'appuie sur des prémisses empiriques bien établies. Pour passer d'une hypothèse scientifique à une théorie scientifique, il faut que l'hypothèse puisse expliquer et/ou prédire des classes de faits indépendants représentés par les boîtes situées dans la partie inférieure du schéma. L'existence d'un modèle analogique n'est pas indispensable mais renforce la théorie scientifique.

(1) Une théorie scientifique repose sur une hypothèse qui elle-même repose sur des faits empiriques généralement bien établis.

(2) Cette hypothèse acquiert le statut de théorie si elle parvient à expliquer des classes de faits indépendants, attendu que par « classes de faits indépendants », on entend non pas des classes de faits qui n'ont aucune interaction entre eux (ce qui reviendrait à avoir une conception extrêmement exigeante des théories scientifiques) mais des classes de faits qui sont produits par des causes indépendantes. Une telle interprétation de la notion de « faits indépendants » suppose une interprétation historique : les faits sont considérés comme causalement indépendants jusqu'à ce que la théorie testée propose une explication qui les unifie sous un même paradigme.

(3) Le dernier critère est un critère facultatif : la théorie gagne en crédibilité si la théorie scientifique s'appuie sur un modèle analogique, c'est-à-dire sur un modèle qui

partage un certain nombre de caractéristiques avec la théorie proposée et dont la validité est déjà bien établie.

4.1.4. Avantages d'une conception déflationniste des théories scientifiques

Ces critères présentent plusieurs avantages. (1) Tout d'abord, ils permettent de distinguer une théorie scientifique d'une simple hypothèse (pour reprendre les termes de Carter, de faire la différence par exemple entre la théorie des germes et de simples opinions sur le rôle des germes dans les maladies infectieuses), sans se prononcer sur les critères d'une « bonne » théorie que nous établirons dans un second temps.

(2) Ensuite, ils sont peu exigeants et expriment une conception déflationniste des théories scientifiques. C'est bien dans cet esprit (et non dans celui du cercle de Vienne) que nous utilisons le terme de théorie et cette démarche est bien conforme au programme « minimalement normatif » que nous nous étions fixés dans l'introduction de ce chapitre : cette définition permet d'identifier un certain nombre de critères qui nous serviront de guide lorsqu'il s'agira d'identifier dans la littérature contemporaine d'éventuelles théories génétiques, tout en étant aussi inclusive que possible.

(3) Cette définition des théories scientifiques fait une place à des notions comme celle de vérifiabilité ou de falsifiabilité, en étant suffisamment ouverte pour pouvoir s'adapter à d'autres sciences (et donc à d'autres méthodes scientifiques) que la physique. Ainsi le premier critère est que l'hypothèse scientifique s'appuie sur des prémisses empiriques bien établies ; tandis qu'un critère facultatif (qui renforce la crédibilité d'une théorie donnée) est l'utilisation d'une théorie analogique. Ces deux critères semblent adaptés à la forme des preuves dans les sciences biomédicales.

(4) Enfin, cette définition nous permet de distinguer entre théorie et explication, et de souligner par là même qu'en faisant l'hypothèse que la généticisation des maladies reflète une théorie génétique, nous formulons l'hypothèse la plus forte et la plus contraignante. On peut ainsi distinguer le concept de « théorie génétique » de celui de simple « explication génétique ». Alors qu'une explication peut se limiter à expliquer un seul fait, le concept de théorie implique l'explication d'une ou plusieurs classes de faits. Ainsi, de la même manière que la théorie de la gravitation universelle de Newton permet d'expliquer à la fois le mouvement des planètes, le mouvement des marées et la chute de la pomme, une théorie déploie une explication

dont la vertu principale est de permettre l'unification de différentes classes de faits sous la même structure théorique. C'est la capacité d'unification d'un ensemble de phénomènes qui distingue la théorie de la simple explication.

Nous avons explicité les raisons pour lesquelles nous faisons l'hypothèse d'une théorie génétique de la maladie et la conception déflationniste des théories scientifiques à laquelle nous adhérons. A ce stade de notre réflexion, on pourrait toutefois nous objecter qu'il existe en philosophie de la médecine un fort scepticisme à l'égard du concept de théorie médicale. Ainsi, un auteur comme Harold Kincaid (Kincaid, 2008) suggère par exemple que parler de théorie médicale revient à importer abusivement un concept généraliste de philosophie des sciences (le concept de théorie scientifique) dans un domaine, la médecine, qui est avant tout orienté vers la thérapeutique (guérir la maladie) et non vers l'acquisition de connaissances scientifiques (connaître la maladie) et que quand bien même elles existeraient, les théories médicales ne seraient pas pour autant utiles. On pourrait également objecter qu'accepter le concept de théorie médicale semble nécessiter de définir préalablement les concepts de santé et de maladie. Enfin, on pourrait objecter qu'accepter le concept de théorie médicale suppose qu'on définisse les maladies ou la maladie comme des espèces naturelles. Afin de ne pas alourdir la lecture de ce chapitre, nous avons choisi de détailler nos objections à ces différentes versions du scepticisme à l'égard du concept de théorie médicale en annexe (Annexe 2 – Prémisses nécessaires pour accepter le concept de théorie médicale). Nous nous contenterons donc ici de présenter nos conclusions. Accepter l'existence de théories médicales nécessite seulement d'accepter trois prémisses peu exigeantes : le concept de théorie médicale a) suppose de reconnaître l'existence d'une science médicale, dont la finalité et le développement ne sont pas entièrement subordonnés et déterminés par la pratique clinique b) est plus facilement accepté par les tenants d'une définition biologique du concept de maladie mais peut être acceptée en principe par quiconque accepte une distinction possible entre fait et valeur en médecine c) peut être adopté, indépendamment de la question de savoir si les maladies individuelles sont des entités naturelles.

À présent que nous avons explicité notre conception déflationniste des théories scientifiques et écarté les objections formulées à l'égard du concept de théorie

médicale, notre objectif est d'une part de définir les conditions nécessaires d'une théorie médicale *a minima*, et d'autre part de déterminer les critères d'une théorie médicale *a maxima* ou idéale, afin de parvenir à décrire le spectre des théories génétiques de la maladie possibles. A ces fins, puisque nous avons défini une théorie comme une explication unifiante, nous examinerons successivement trois questions : qu'est-ce qu'une théorie médicale explique ? (Section 4.2.) Comment une théorie médicale explique-t-elle ? (Section 4.3.) Jusqu'où une théorie doit-elle expliquer ? (Section 4.4.).

4.2. Qu'est-ce qu'une théorie médicale explique ?

Afin de définir ce qu'est une théorie médicale, il faut d'abord définir son objet, c'est-à-dire ce qu'elle est supposée expliquer. Dans cette section, nous proposons une série de distinctions conceptuelles permettant d'identifier un objet spécifique des théories médicales (par rapport aux théories biologiques), d'analyser la variété des niveaux d'explication en médecine et de rendre compte de la diversité des cibles explicatives dans l'explication d'une maladie.

4.2.1. L'objet des théories médicales est l'explication de la maladie

Historiquement, les théories médicales ont été conceptualisées sur le modèle des théories biologiques, au point que certains auteurs, comme Kenneth Schaffner par exemple, distinguent rarement les deux et parlent de « théories biomédicales ». Si Kenneth Schaffner a sûrement proposé une des caractérisations les plus précises des théories biomédicales (Schaffner, 1980, 1986, 1993, 1998, 2006, 2008, 2011), nous voudrions insister sur un certain nombre de caractéristiques très générales sur lesquelles s'accordent des auteurs aussi différents que Schaffner, Thompson (Thompson, 2011 ; Thompson et Gifford, 2011) et Kazem Sadegh-Zadeh (Sadegh-Zadeh, 2000, 2008, 2011) et avec lesquelles la conception des théories médicales que nous développons est compatible.

Ainsi (1), les théories biomédicales ne sont pas universelles : Schaffner emploie le terme de « moyenne portée » pour décrire ce phénomène et différencier le niveau

de généralité des théories biomédicales de celui des théories physiques. En effet, les théories biomédicales s'intéressent au vivant, et sont donc confrontées en permanence aux phénomènes de variation interindividuelle, de plasticité, d'hétérogénéité phénotypique, de pléiotropie et de réalisation multiple, qui rendent extrêmement difficile la formulation de lois générales ou d'affirmations universelles. Cela ne signifie pas que les théories biomédicales sont cantonnées à la description du particulier : le caractère contrefactuel des mécanismes décrits par la théorie suffit à assurer un niveau de généralité acceptable.

(2) Parce que les théories biomédicales sont orientées vers un objet (le vivant) dont les caractéristiques principales sont la variation interindividuelle et l'existence de processus temporels intervenant simultanément à plusieurs niveaux (cellulaire, organique, systémique), les théories biomédicales utilisent des modèles qui exemplifient un certain nombre de traits communs aux membres d'une même classe de modèles – c'est la ressemblance avec ce modèle principal qui assure l'appartenance de modèles différents à une même classe de modèles. Schaffner utilise le concept de « modèle prototypique » pour décrire ce phénomène, Kazem Sadegh-Zadeh parle également de « prototype de la maladie » (Sadegh-Zadeh, 2008).

(3) Enfin, les théories biomédicales sont fondamentalement transdisciplinaires : elles peuvent parfois être formalisées avec des modèles mathématiques et font appel à tous les types de connaissances nécessaires à la découverte des mécanismes du vivant (la biologie moléculaire, la biologie cellulaire, la biochimie, la théorie de l'évolution, etc.).

Si nous reconnaissons que ces caractéristiques sont en effet communes aux théories biomédicales, nous nous opposons à l'absence de distinction entre théories biologiques et théories médicales, ce que semble soutenir Schaffner en parlant régulièrement de « théories biomédicales » et que soutient explicitement Paul Thompson dans une série d'articles (Thompson, 2010, 2011 ; Thompson et Gifford, 2011). Dans son analyse, Thompson insiste sur la diversité des théories médicales, en s'appuyant sur trois exemples : le modèle insuline/glucose de Bolie, le modèle de la régulation du cycle menstruel de la femme et le modèle de la description d'une épidémie de peste.

Le modèle insuline-glucose est un modèle mathématique, emprunté à la physiologie, décrivant la relation entre les concentrations d'insuline et de glucose dans un organisme normal. Ce n'est cependant pas une explication du diabète : ce modèle ne relie pas les symptômes du diabète à une résistance ou à une absence d'insuline, il n'explique pas non plus la manière dont se met en place une insulino-résistance au cours du diabète de type 2, ou la façon dont les îlots de Langerhans sont détruits au cours du diabète de type 1. D'ailleurs, il est intéressant de noter que ce type de modèle mathématique n'est jamais utilisé en pratique clinique : même lorsqu'il s'agit d'instaurer un traitement à l'insuline pour un patient diabétique, le traitement est instauré sur la base empirique d'une augmentation progressive des doses d'insuline jusqu'à atteindre un équilibre souhaité et qui est tout aussi empiriquement constaté par le contrôle des glycémies du patient à intervalles réguliers sur plusieurs jours.

Le second modèle qu'analyse Thompson est le modèle du cycle menstruel, c'est-à-dire le modèle de la régulation du cycle menstruel par rétrocontrôle des hormones secrétées par les ovaires (FSH et LH) et des hormones secrétées par l'hypophyse (GnRh, LnRH). Un tel modèle n'est l'explication d'aucune maladie en soi, bien qu'il puisse intervenir dans l'explication d'autres maladies. Thompson lui-même reconnaît que la seule intervention directe qu'on puisse déduire du modèle du cycle menstruel dans le cadre médical est la contraception hormonale (Thompson et Gifford, 2011, 122) dont on pourrait discuter a) qu'elle constitue à proprement parler un traitement médical (plutôt qu'une intervention « de confort ») b) et *a fortiori* la portée emblématique en termes de thérapeutique.

Le troisième exemple que convoque Thompson concerne l'explication épidémiologique de la Peste Noire. Pour cet exemple, Thompson développe le modèle de Reed-Frost, un modèle proposé en 1928, pour décrire la transmission des maladies infectieuses et modéliser les épidémies. Thompson montre que ce modèle a été utilisé pour déterminer l'étiologie infectieuse de la Peste Noire : en se fondant sur les observations de l'époque et en comparant la vitesse de propagation de l'épidémie de la peste noire à la peste bubonique, des chercheurs ont ainsi pu démontrer que la cause de la peste noire n'était probablement pas l'agent de la peste bubonique, mais un autre agent infectieux (Duncan et Scott, 2005). Peut-on considérer que le modèle de Reed-Frost soit une théorie médicale ? Par opposition aux exemples précédents, le

modèle de Reed-Frost est certainement utile pour expliquer directement une maladie. Il constitue toutefois moins une explication de la maladie qu'un modèle général d'épidémie utilisé non pas pour inférer la cause de la peste noire, mais pour écarter une mauvaise hypothèse causale, dans le cas d'un fait unique : l'épidémie de la Peste Noire. De ce point de vue, le modèle de Reed-Frost participe certainement à l'explication de la Peste Noire, mais il ne constitue pas une théorie qui explique une classe de faits indépendants – seulement une forme d'explication (l'explication épidémiologique) d'un fait ponctuel[27].

Nous soutenons ainsi qu'aucun des exemples que prend Thompson ne constitue une théorie médicale à proprement parler, car aucune de ces théories ne constitue une explication des maladies, contrairement à ce que l'usage historique considère généralement être des théories médicales comme la théorie des germes, par exemple, qui est une explication générale des maladies infectieuses. Précisons d'ailleurs que cette thèse de Paul Thompson a un objectif bien précis : identifier les théories médicales à des théories biologiques est un moyen de leur donner une légitimité et de distinguer fortement les théories médicales qui appartiennent à la science médicale de la médecine des preuves (EBM), dont Thompson soutient qu'elle appartient avant tout à la pratique clinique et qu'elle ne peut à ce titre constituer un fondement pour la science médicale. S'il nous semble donc que Paul Thompson a raison d'insister sur les similitudes entre théories biologiques et théories médicales et sur la diversité des théories médicales, les exemples qu'il choisit démontrent une confusion entre le concept de théorie médicale et les théories biologiques ou mathématiques dont se sert ponctuellement la médecine. Il y a bien une spécificité des théories biologiques par rapport aux théories médicales et c'est justement l'objet de la théorie : les théories biologiques expliquent toutes sortes de phénomènes du vivant, mais les théories médicales expliquent les maladies.

[27] Nous reviendrons sur ce point un peu plus loin au moment de justifier pourquoi l'explication de faits individuels ne peut être le socle d'une théorie médicale (section 4.2.3).

4.2.2. Différents niveaux de théorie médicale : théorie de la maladie, d'une classe de maladie et d'une maladie

Nous avons donc établi le premier critère (faussement trivial) d'une théorie médicale : les théories médicales sont avant tout des explications de maladie. La difficulté est que le concept de maladie est lui-même ambigu et c'est pourquoi il nous semble utile de nous appuyer ici sur un ensemble de distinctions conceptuelles proposées par Kazem Sadegh-Zadeh entre le concept de maladie, les classes de maladie et les maladies individuelles d'une part, et le concept de maladie, la maladie-type et la maladie-instance d'autre part :

- *la* maladie (*the disease*) désigne la catégorie « maladie »
- *une* maladie individuelle (*individual disease*) est un membre de la catégorie générale « maladie » et désigne un des 50 000 « prédicats nosologiques » que nous utilisons dans la pratique quotidienne de la médecine pour rattacher un ensemble de symptômes à un diagnostic (Sadegh-Zadeh, 2011, p. 154–155). On utilise deux concepts, la maladie-type (*type-disease*) et la maladie-instance (*token-disease*) pour caractériser deux usages différents des prédicats nosologiques ou maladies individuelles
    - La *maladie-type* désigne un prédicat nosologique (exemple : la tuberculose) en tant que classe de phénomènes
    - la *maladie-instance* désigne la maladie telle qu'elle est incarnée par le patient malade (exemple : la tuberculose de Marie) (Sadegh-Zadeh, 2011, p. 155–156).
- *une classe de maladie* (*disease class*) *:* les prédicats nosologiques ou maladies individuelles peuvent eux-mêmes être regroupés en plusieurs classes de maladie, telles que « les maladies infectieuses », ou « les maladies cardiovasculaires » (Sadegh-Zadeh, 2011, p. 151–152).

Quel rapport entre ces différentes distinctions conceptuelles et le concept de théorie médicale ? Franchissant un pas supplémentaire par rapport au propos de Kazem Sadegh-Zadeh, nous affirmons qu'aux différents types d'objets-maladies que nous avons identifiés, correspondent différents niveaux de théorie médicale, comme le montre l'histoire des théories médicales que propose Thagard. En effet, dans sa

description de l'histoire des théories médicales, Paul Thagard identifie une coupure fondamentale entre les théories médicales antiques (comme la théorie humorale par exemple) et les théories scientifiques modernes (qui seraient nées à partir du milieu du dix-neuvième siècle et de l'avènement de la théorie des germes). (Thagard, 2006, p. 46)

Parmi les théories médicales antiques, Thagard prend pour exemple la théorie humorale hippocratico-galénique, la théorie de la médecine traditionnelle chinoise et la théorie de la médecine traditionnelle indienne. Chacune de ces théories propose une explication de *la maladie* en général, la maladie étant définie comme une forme de déséquilibre. Ainsi, dans la théorie humorale, la maladie naît d'un déséquilibre entre quatre humeurs du corps. C'est en fonction de cette définition de *la maladie* en général que toutes les maladies-types sont expliquées. À partir de la médecine scientifique moderne, que Thagard date de la théorie des germes, il n'y a plus de théorie générale de la maladie mais uniquement des théories de classes de maladie. Ainsi, la théorie des germes explique la classe des maladies infectieuses. Chaque maladie individuelle est ensuite expliquée en fonction de sa classe : ainsi, chaque maladie infectieuse individuelle (tuberculose, SIDA, etc.) sera expliquée par l'identification de son pathogène spécifique (Thagard, 2005, p. 48–49).

Thagard représente ainsi notre théorie médicale contemporaine (Figure 14) : c'est un ensemble hétérogène de théories de classes de la maladie. Il est intéressant d'ailleurs de remarquer, que conformément à l'usage habituel que nous contestons, les maladies génétiques apparaissent comme une classe de maladie parmi d'autres dans cette théorie médicale contemporaine.

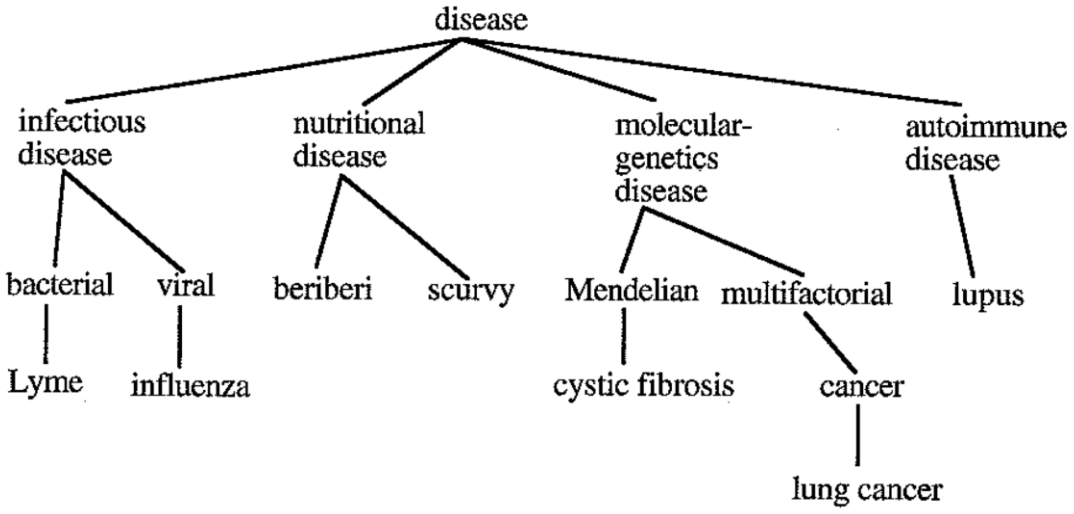


Figure 14 : Représentation schématique de l'organisation hiérarchique des explications des maladies dans la théorie médicale contemporaine avec des exemples de maladies individuelles (Thagard, 2000, p. 35)

Nous nous proposons d'utiliser cette distinction entre théorie antique et théorie moderne pour construire une typologie du niveau de l'explication des théories médicales. Si on parle en termes de « niveaux d'explication », c'est-à-dire en l'occurrence en termes de « niveaux de généralité », on peut distinguer trois niveaux fondamentaux sur le schéma de Thagard (Figure 14) :

- Le niveau le plus haut est celui d'une *théorie médicale générale*, qui est supposée pouvoir valoir pour toutes les maladies
- Le niveau intermédiaire est le niveau de *la théorie médicale régionale* qui vaut pour une classe spécifique de maladies. Remarquons d'ailleurs qu'il y a en fait sur ce schéma deux niveaux intermédiaires : par exemple, en ce qui concerne la maladie infectieuse, on peut distinguer un niveau intermédiaire élevé (la classe des maladies infectieuses) qui se divise en sous-classes de maladie (la classe des maladies bactériennes, la classe des maladies virales, etc.). Remarquons également qu'il n'y a pas de principe unificateur dans la façon dont sont délimitées les classes de maladie : ainsi, les maladies infectieuses correspondent à une classe étiologique (c'est l'ensemble des maladies dont l'une des causes est l'infection par un pathogène), alors que les maladies

cardiovasculaires correspondent à une classe anatomo-fonctionnelle (l'ensemble des maladies du cœur et des vaisseaux).

- Le niveau inférieur est le niveau de la *maladie individuelle*.

Par ailleurs, si on applique la distinction entre théorie antique et théorie médicale contemporaine, on peut distinguer deux types de théorie générale (correspondant au niveau supérieur) : dans les théories antiques, la théorie médicale générale est une *théorie de la maladie*, qui repose sur une définition générale de la maladie à partir de laquelle l'explication de chaque maladie-type peut être individualisée. Dans notre théorie médicale contemporaine, la théorie médicale générale est une *théorie des maladies* : c'est un ensemble de théories régionales, qui reposent sur l'identification d'une étiologie, d'un mécanisme ou d'une localisation anatomo-clinique commun à une classe de maladie[28].

4.2.3. Quatre cibles explicatives communes : causes, symptômes, évolution, traitement

Nous avons insisté sur l'existence de différents objets-maladies, ce qui nous conduit à affirmer l'existence de différents niveaux de théorie médicale. On peut néanmoins se demander s'il y a des points communs entre ces différents niveaux de théorie. Suivant ici Paul Thagard, nous considérons qu'expliquer une maladie, c'est expliquer quatre classes de faits, que nous appellerons ici des « cibles explicatives » et qu'on peut retrouver ces quatre cibles explicatives pour tous les objets-maladies et donc pour tous les niveaux de théorie médicale. Ces quatre cibles explicatives sont les causes, les symptômes, l'évolution et le traitement. Thagard représente les relations qui existent entre ces différentes cibles explicatives dans le schéma suivant (Figure 15) :

[28] Cette typologie peut paraître triviale, toutes ces composantes ne seraient cependant pas aisément acceptées par tous les philosophes de la médecine, si on se souvient du débat sur la nature des maladies, définies ou pas comme des espèces naturelles, que nous avons très brièvement évoqué précédemment et qui est détaillé en annexe (Annexe 2). C'est le concept de théorie de la maladie (au sens antique du terme) qui poserait en particulier le plus de difficultés car il semble supposer qu'on considère que la maladie est une espèce naturelle, au-delà de la diversité apparente des maladies individuelles. Nous montrons cependant (Annexe 2) qu'on ne peut pas démontrer *a priori* que la maladie ne pourra jamais être une espèce naturelle. Il s'ensuit qu'une théorie de la maladie est en principe possible.

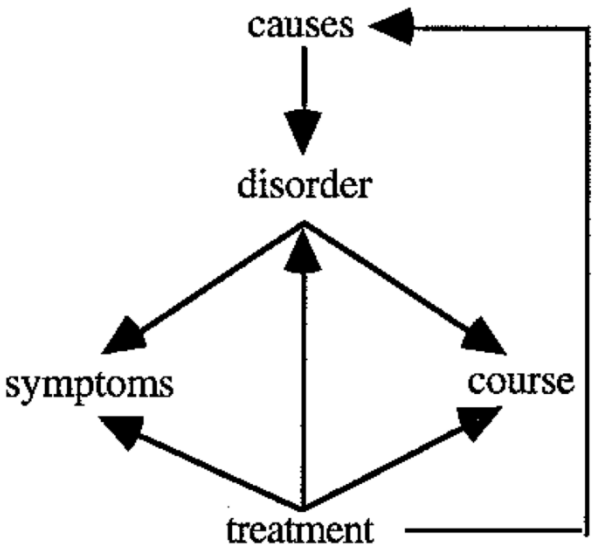


Figure 15 : Structure causale du concept de maladie (Thagard, 2000, p. 23). Un faisceau de causes entraine l'apparition de la maladie, ses symptômes et son évolution. Le traitement peut agir sur les symptômes, sur l'évolution de la maladie ou directement sur ses causes.

Selon Thagard, chaque maladie peut être représentée par un schéma explicitant les relations causales identifiées dans la Figure 15. Les symptômes désignent l'ensemble des manifestations observables (cliniquement ou infracliniquement, ce qui suppose l'intervention de tests médicaux) d'une maladie donnée (maladie $\rightarrow$ symptômes). Ces symptômes peuvent apparaître sur une certaine période de temps et se développer de différentes manières, ce qu'on appelle l'évolution ou l'histoire de la maladie (maladie $\rightarrow$ évolution). Les symptômes sont eux-mêmes le produit des facteurs causaux qui se sont exercés sur le patient (causes $\rightarrow$ maladie). L'objectif du traitement est d'influencer à la fois les symptômes (traitement $\rightarrow$ symptômes) et l'évolution de la maladie (traitement $\rightarrow$ évolution), la plupart du temps en intervenant sur les causes de la maladie responsables de l'apparition des symptômes (traitement $\rightarrow$ causes).

Si nous sommes globalement d'accord avec la description générale de Thagard, ajoutons simplement que le concept de « causes de la maladie » est un concept pour le moins ambigu, mais que nous clarifierons dans la suite de ce chapitre (Section 4.3), en exposant la façon dont différents types d'explications d'une maladie font appel à différents types de causes.

Ces quatre cibles explicatives de la maladie sont communes en un double sens : (1) parce qu'elles interviennent aussi bien dans l'explication de la maladie-type que de la maladie-instance (2) parce qu'elles interviennent dans des théories de différents niveaux.

(1) Ce schéma s'applique à la maladie individuelle en tant qu'elle est une maladie-type et en tant qu'elle est une maladie-instance. Lorsque l'on cherche à expliquer une maladie type, on veut recenser l'ensemble des causes qui peuvent expliquer une maladie donnée, l'ensemble des symptômes possibles et les différentes évolutions possibles de la maladie. C'est ainsi que les manuels de médecine présentent spontanément l'ensemble des facteurs de risque et des mécanismes physiopathologiques impliqués dans la maladie et recensent de façon probabiliste les différentes présentations cliniques d'une même maladie, les circonstances de sa découverte. Lorsque le clinicien se trouve devant un malade, il se trouve confronté au chemin inverse : à partir de l'ensemble des symptômes qu'il observe, il doit poser un diagnostic de la maladie. Une fois la maladie identifiée, en fonction des caractéristiques de l'histoire médicale de son patient, il envisage alors quelle est l'évolution probable de la maladie et quels sont les traitements possibles. L'ensemble des causes n'est pas forcément convoqué dans cette explication : au contraire, il s'agit de sélectionner quelles sont les causes pertinentes qui sont susceptibles d'expliquer pour ce patient en particulier l'apparition de la maladie et les symptômes qu'il présente. C'est ainsi que Thagard parle d' « instanciation d'un réseau causal » pour passer des causes possibles de la maladie (les causes de la maladie-type) aux causes actuelles de la maladie de ce patient (les causes de la maladie-instance). Par exemple, Thagard représente ainsi le schéma des causes de l'ulcère (Figure 16) :

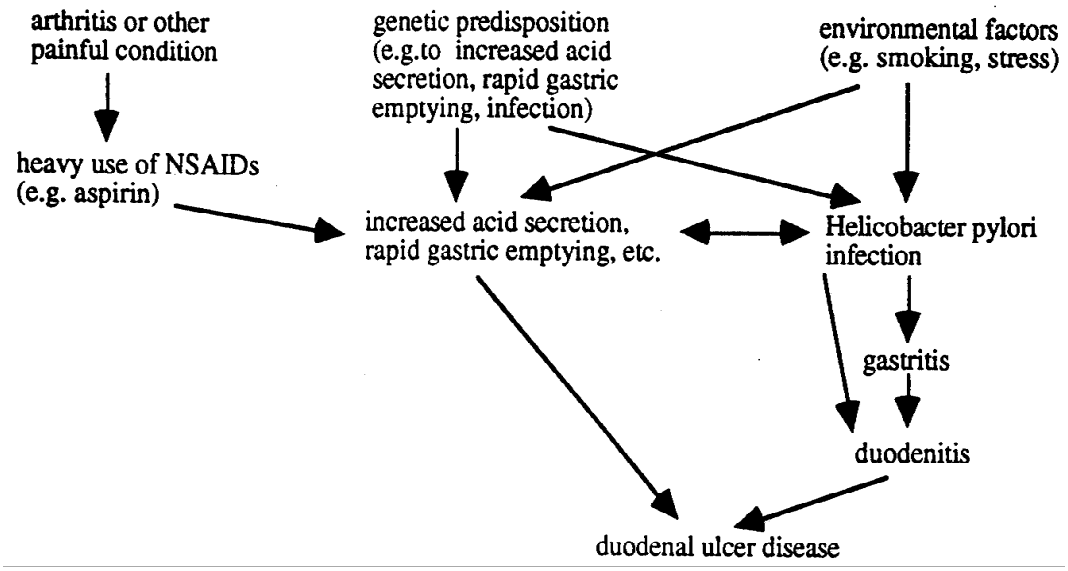

arthritis or other painful condition
genetic predisposition (e.g.to increased acid secretion, rapid gastric emptying, infection)
environmental factors (e.g. smoking, stress)
heavy use of NSAIDs (e.g. aspirin)
increased acid secretion, rapid gastric emptying, etc.
Helicobacter pylori infection
gastritis
duodenitis
duodenal ulcer disease

Figure 16 : Réseau causal général de l'ulcère duodénal (Thagard, 1998, 74)

Comme on peut le remarquer le schéma causal est assez compliqué : l'ulcère est une maladie multifactorielle au sens où il peut être expliqué par des plusieurs causes qui peuvent être indépendantes mais sont souvent intriquées dans différentes relations. Mais l'ensemble de ces causes ne seront pas retrouvées dans l'explication de l'ulcère de Julia : en fonction des différents tests diagnostiques et de l'histoire médicale de la patiente, on va pouvoir instancier le schéma d'explication des causes de l'ulcère en tant que maladie-type à l'explication de l'ulcère de Julia en tant que maladie-instance (Thagard, 1998, 74).

Thagard s'intéresse ici à l'instanciation du réseau des *causes* de l'ulcère de Julia mais c'est la même démarche d'instanciation qui permet de reconnaître l'ulcère de Julia même si elle ne présente pas tous les *symptômes* de la maladie-type et c'est la même démarche d'instanciation qui va s'appliquer pour prédire l'*évolution* de la maladie de Julia et pour déterminer quel est le *traitement* à mettre en place pour le traitement de l'ulcère de Julia. On retrouve donc nos quatre catégories de cibles explicatives dans l'explication de la maladie individuelle en tant que maladie-type et en tant que maladie-instance, même si le passage de l'explication de la maladie-type à la maladie-instance suppose une instanciation.

(2) Ce schéma peut être instancié dans des théories de différents niveaux (théorie des germes et théorie hippocratique). Autrement dit, on retrouve ces quatre mêmes catégories de cibles explicatives, quel que soit le niveau d'objet-maladie que l'on cherche à expliquer. Ainsi, Thagard montre dans les deux schémas ci-dessous (Figure 17) comment le concept de maladie est expliqué dans la théorie hippocratico-galénique, exemple de ce que nous avons appelé une théorie générale de la maladie (Figure 17, à gauche) et comment le concept de maladie est expliqué dans la théorie des maladies infectieuses, exemple de ce que nous avons appelé une théorie régionale d'une classe de maladie (Figure 17, à droite) :

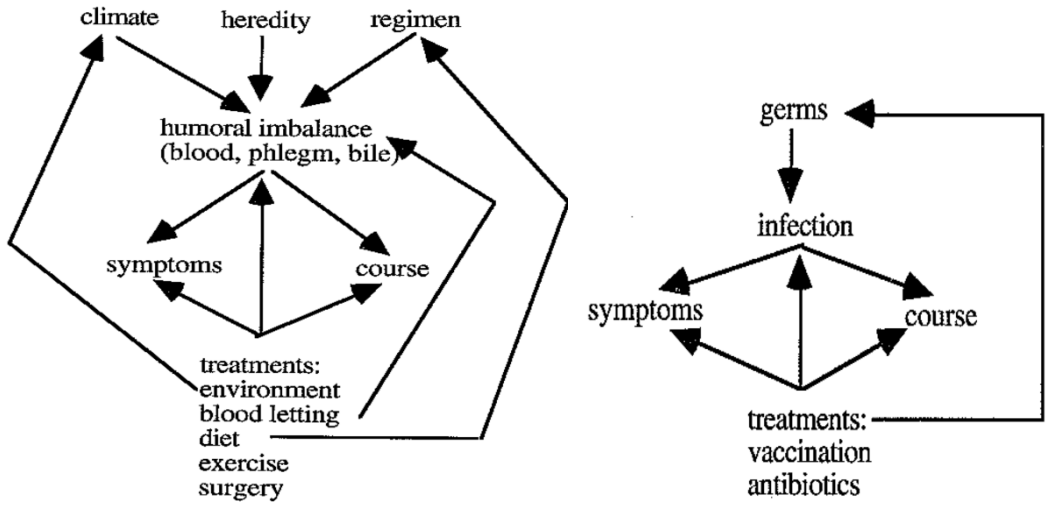


Figure 17 : Comparaison des structures causales de la maladie dans la théorie hippocratique (à gauche) (Thagard, 2000, p. 23) et dans la théorie des germes (à droite) (Thagard, 1998, p. 25)

Que l'on soit dans une théorie générale de la maladie ou dans une théorie régionale d'une classe de maladie, on retrouve donc nos quatre cibles explicatives : causes, symptômes, évolution, traitement. Et chaque schéma explicatif peut être instancié pour expliquer une maladie-type. Dans la théorie hippocratique, on expliquera l'épilepsie comme un ensemble de symptômes (crises épileptiques) causés par une excès de phlegme et qui doit être traitée en débarrassant le patient de son excès de phlegme. Ce schéma peut alors être instancié à son tour pour expliquer l'épilepsie de Marie, en identifiant quels facteurs (climats, hérédité, régime) ont produit le déséquilibre humoral et permettant une adaptation du traitement en fonction des symptômes présentés par Marie, de l'évolution de sa maladie dans le

temps et des facteurs spécifiques à son cas individuel qui ont été identifiés comme la cause de son déséquilibre humoral. Dans la théorie des germes, on expliquera la maladie infectieuse comme un ensemble de maladies produites par des germes spécifiques pour chaque maladie et dont le traitement dépend de l'identification du germe. De la même manière, on pourra instancier ce schéma pour expliquer la tuberculose en général et la tuberculose de Marie en particulier.

Dans cette section, notre objectif était d'identifier la spécificité de l'objet des théories médicales. Si les théories médicales partagent de nombreuses caractéristiques avec les théories biologiques, il nous semble que la spécificité des théories médicales réside dans son objet : l'explication des maladies. Nous avons montré que l'on pouvait distinguer différents niveaux objets-maladies, selon qu'on se place du point de vue de la maladie en général, d'une classe de maladie ou de la maladie individuelle, qui est elle-même comprise dans sa dimension générale comme une maladie-type et dans sa dimension particulière comme une maladie-instance. A cette typologie des différents objets-maladies correspond une typologie des différentes théories médicales possibles, selon qu'on se place au niveau de l'explication de toutes les maladies, d'une classe de maladie, ou de la maladie individuelle et selon qu'on admette ou pas une définition générale de la maladie. Mais, quel que soit le niveau de l'objet-maladie que l'on cherche à expliquer et que l'on se trouve dans une théorie d'une maladie individuelle, dans une théorie d'une classe de maladie ou dans une théorie générale de la maladie, nous avons mis en évidence quatre cibles explicatives communes : les causes, les symptômes, l'évolution et le traitement. À présent que nous avons caractérisé ce que les théories médicales expliquent, il nous faut dire comment elles expliquent.

4.3. Comment une théorie médicale explique-t-elle ?

4.3.1. Le pluralisme explicatif en médecine

Il existe différents types d'explications en médecine et chacune a sa propre spécificité et sa propre légitimité. Maël Lemoine identifie ainsi tout un ensemble d'explications qui ont des objets différents et qui font appel à des méthodes et à des

disciplines différentes, mais qui appartiennent toutes au domaine de la science médicale (Lemoine, 2011) : explication biologique, explication pharmacologique, explication épidémiologique, explication clinique…

Ce pluralisme explicatif en médecine peut être compris comme un pluralisme explicatif méthodologique, qui dépend de nos intérêts épistémiques et qui a trait à la variété des questions que l'on peut poser, et/ou comme un pluralisme ontologique ou de principe, tenant à la nature des phénomènes que l'on cherche à expliquer. Les deux approches ne sont pas contradictoires et peuvent être tenues ensemble, mais l'accent est différent. Le pluralisme méthodologique met l'accent sur la diversité des questions en médecine (se demander pourquoi cet individu est malade est différent de se demander pourquoi cette population est malade) et suppose une approche pragmatique des explications, inspirée de Van Fraassen (Van Fraassen, 1980). C'est ainsi que Maël Lemoine met en avant l'interdisciplinarité constitutive de la médecine pour expliquer le pluralisme explicatif :

> « La diversité des champs de la connaissance où les explications médicales de différents types ont cours ne peut assurément ni reproduire, ni ignorer la délimitation des spécialités ou des disciplines médicales. Les spécialités peuvent en effet être classées par organes ou tissus (néphrologie, dermatologie, cardiologie, gastro-entérologie, neurologie, hématologie), par techniques (radiologie, épidémiologie), par moyens thérapeutiques (pharmacologie, chirurgie), par maladies (oncologie, diabétologie, allergologie), par aspects étiologiques des maladies (virologie, bactériologie), par terrains (gériatrie, gynécologie, pédiatrie), par emprunts à des disciplines non médicales (biostatistiques, biophysiques, génétique médicale, pharmacocinétique, biologie cellulaire et moléculaire). Il est probable que des types d'explications sont communs à plusieurs spécialités et qu'au sein de certaines d'entre elles du moins, on peut recourir à des explications de types différents. » (Lemoine, 2011, p. 26)

Le pluralisme ontologique affirme un anti-réductionnisme fondamental : c'est parce que les phénomènes cliniques, par exemple, ne sont pas réductibles aux phénomènes biologiques que l'on est fondé à envisager légitimement plusieurs types d'explications correspondant à plusieurs niveaux d'explication d'un phénomène. C'est

le type de pluralisme ontologique défendu par Sandra Mitchell par exemple (Mitchell, 2003 ; Mitchell et Dietrich, 2006). Maël Lemoine est plus prudent à ce sujet, considérant que l'irréductibilité des différents types d'explication médicale ne peut pas être déclarée par principe mais que l'existence de différents modes explicatifs suggère une irréductibilité de fait :

> « La désunité des explications d'un même phénomène est produite par un effet de superposition : il s'agit du même phénomène, mais chaque explication est « irréductible » aux autres explications. Généralement, on entend par « irréductibilité » l'affirmation de l'autonomie de principe d'une discipline (ou d'un type d'explication), et par « réductibilité », l'affirmation de sa dérivabilité (de principe) à partir d'une autre discipline ou d'un autre type d'explication. Par « désunité », on entend souligner que « l'irréductibilité » des différents types d'explications de la science médicale les uns par rapport aux autres n'est pas de principe, mais de fait : il ne s'agit pas d'affirmer qu'ils ne pourront pas être réduits ou éliminés, mais de ne pas affirmer qu'ils le seront ou ne le seront pas. » (Lemoine, 2011, p. 188)

C'est sur cette double affirmation (pluralisme méthodologique et pluralisme ontologique de fait) que nous nous appuierons dans la suite de cet exposé. Trois questions nous paraissent ainsi essentielles lorsqu'on cherche à expliquer une maladie, et nous allons en montrer l'importance avec l'exemple de la drépanocytose. Ces trois grandes questions sont : pourquoi cette population est-elle malade? Comment la maladie advient-elle ? Pourquoi cet individu est-il malade ? A chacune de ces questions correspondent un ou plusieurs types d'explications.

(1) La question « pourquoi cette population est-elle malade ? » peut mener à deux types d'explication, (1a) selon que l'on s'intéresse aux facteurs de risque de la maladie – auquel cas on développera une *explication épidémiologique* de la maladie, ou (1b) selon que l'on s'intéresse aux causes évolutionnaires de la maladie – auquel cas on développera une *explication évolutionnaire* de la maladie (1b). (2) La question « pourquoi cette maladie advient-elle ? » amène à développer une explication *physiopathologique* qui s'intéresse aux mécanismes biologiques qui expliquent l'apparition de la maladie et l'ensemble des symptômes mais aussi l'évolution de la maladie dans le temps. Enfin (3), la question « pourquoi cet individu est-il malade ? »

amène à développer une explication clinique de la maladie : il s'agit d'instancier à partir des caractéristiques du patient, de ses antécédents personnels et familiaux, de son mode de vie, quelles sont les causes de la maladie qu'il présente et de prédire comment sa maladie va évoluer.

Une description exhaustive de ces différents types d'explication dépasserait très largement les limites de notre travail et a déjà été proposée ailleurs[29]. Mais dans la mesure où nous ferons référence à ces types d'explications dans la suite de notre travail (partie 3), il est cependant nécessaire de présenter brièvement leurs caractéristiques les plus saillantes et les rapports qu'elles entretiennent entre elles. En particulier, nous voulons mettre en exergue le fait que ces trois explications (1) sont toutes susceptibles de concerner l'explication de toutes les maladies (2) sont toutes susceptibles d'expliquer toutes les cibles explicatives d'une maladie et pas seulement l'explication des causes d'une maladie (3) ont toutes une nature contrastive[30], c'est-à-dire qu'elles tirent leur pouvoir explicatif de la comparaison d'un même phénomène (les causes, les symptômes, l'évolution ou le traitement d'une maladie) dans des populations différentes et (4) ont toutes une spécificité qui fait qu'elles ne sont ni substituables ni remplaçables les unes par rapport aux autres mais complémentaires – ce qui est intrinsèquement lié au pluralisme explicatif que nous défendons.

4.3.2. L'explication épidémiologique

L'épidémiologie peut être définie comme *« l'étude des déterminants de la maladie et des autres états de santé des populations humaines au moyen de comparaison de groupes dans l'objectif d'améliorer la santé des populations »* (Broadbent, 2013, p. 1). L'explication épidémiologique de la maladie est ici définie

[29] Pour une description des explications épidémiologiques, on se référera en priorité aux travaux d'Alex Broadbent (Broadbent, 2013). Pour une description des explications physiopathologique et clinique, on se référera aux travaux de Maël Lemoine (Lemoine, 2011).

[30] Nous entendons le terme « nature contrastive » au sens large d'explication utilisant des classes de contraste. Plus précisément, nous soutenons que si toutes ces explications (épidémiologique, évolutionnaire, physiopathologique) tirent leur pouvoir explicatif de l'utilisation des classes de contraste, cette affirmation prend des sens différents en fonction du type d'explication considérée. Ainsi, l'explication épidémiologique, intrinsèquement populationnelle-statistique tire son pouvoir explicatif de l'utilisation de classes de contrastes, formés d'individus réels. L'explication physiopathologique tire son pouvoir explicatif du contraste entre des mécanismes normaux et des mécanismes physiopathologiques. L'explication évolutionnaire (au sens darwinien du terme) tire son pouvoir explicatif du contraste entre l'ensemble des contraintes qui s'exercent sur l'homme moderne et les contraintes exercés sur l'homme du Pléistocène.

comme une explication populationnelle de la maladie. En effet, un énoncé épidémiologique n'est jamais supposé s'appliquer à un individu particulier mais aux groupes de populations qui sont étudiés. Une explication épidémiologique ne peut affirmer ni un énoncé du type « Fumer cause le cancer du poumon » ni un énoncé du type « La cause du cancer du poumon de Julia est le tabac ». Un énoncé épidémiologique sera plutôt typiquement de la forme : « Le risque relatif de développer au cours de leur vie un cancer du poumon est de vingt pour les individus qui fument un paquet de cigarettes par jour par rapport aux non-fumeurs ». L'épidémiologie propose donc toujours des énoncés qui portent sur une population en ayant utilisé une autre population (d'individus sains, d'individus malades mais exposés à des facteurs de risque différents, d'individus malades mais exposés à un traitement différent) comme classe de contraste. La question du choix des populations que l'on compare dans une étude, comme la question de la généralisation des conclusions d'une étude portant sur une population à d'autres populations ou à la maladie d'intérêt en tant qu'elle est une maladie-type, sont d'ailleurs des questions centrales pour interpréter les résultats d'une enquête épidémiologique et passer d'une simple association statistique entre deux variables (par exemple le tabac et le cancer du poumon) en comparant deux populations particulières à une interprétation causale de cette corrélation.

Il s'ensuit que l'explication épidémiologique utilise pour principale méthode des comparaisons entre des groupes de populations. On distingue généralement trois types d'études en épidémiologie : les études de cas-témoins[31], les études de cohorte[32] et les essais randomisés contrôlés[33]. C'est précisément la variété des types d'études

[31] Les études de cas-témoins cherchent à déterminer rétrospectivement les facteurs de risque d'une maladie en comparant une population d'individus malades (les cas) à une population d'individus sains (les témoins).

[32] Les études de cohorte sont des études prospectives. Partant de l'hypothèse que l'exposition à tel facteur d'intérêt est un facteur de risque pour le développement d'une maladie, on constitue deux cohortes d'individus exposés et non exposés à ce facteur et on observe pendant plusieurs années l'incidence de la maladie dans les deux groupes.

[33] Les essais randomisés contrôlés différent des deux types d'étude précédemment exposés parce qu'ils consistent non pas à déterminer les facteurs de risque de développer une maladie mais à évaluer l'efficacité thérapeutique d'un ou de plusieurs traitements. Dans ce type d'étude, des individus malades sont répartis aléatoirement en deux groupes dont les caractéristiques sont distribuées de façon similaire dans chaque groupe. L'un des deux groupes recevra le traitement dont on veut tester l'efficacité, tandis que l'autre groupe recevra soit un traitement-placebo, soit le meilleur traitement connu à ce jour. On évalue ensuite l'efficacité comparée des deux traitements.

épidémiologiques qui permet à l'explication épidémiologique d'intervenir plusieurs étapes de l'explication d'une maladie : à la fois dans la détermination des causes de la maladie mais aussi dans l'étude des symptômes d'une maladie, dans l'étude des facteurs de pronostic d'une maladie et dans l'évaluation de l'efficacité des traitements. L'explication épidémiologique peut ainsi évaluer l'association statistique et le lien causal qui existe entre un facteur d'exposition et une maladie, un facteur d'exposition et le développement d'une forme clinique spécifique d'une maladie (c'est-à-dire d'un ensemble de symptômes particuliers), un facteur d'exposition et le pronostic d'une maladie (c'est-à-dire la façon dont la maladie va évoluer en fonction de certains facteurs de risque), un type de traitement employé et un stade de la maladie.

C'est parce que l'explication épidémiologique est ainsi caractérisée (explication populationnelle contrastive, différentes méthodes de comparaison de groupes, différentes cibles explicatives) qu'il n'existe pas une, mais des explications épidémiologiques d'une maladie ou d'une classe de maladie donnée. Ainsi, dans le cas de la drépanocytose, il n'y a pas une explication épidémiologique de la drépanocytose mais plusieurs explications possibles selon que l'on s'intéresse aux causes déclenchantes de la maladie comme par exemple la distribution des variants alléliques de la drépanocytose (Serjeant, 1981), au lien entre certaines caractéristiques d'une population et certains symptômes, comme les facteurs de risque de mortalité dans la drépanocytose ou les facteurs déclencheurs des syndromes aigus thoraciques (Platt *et al.*, 1994 ; Castro *et al*., 1994) ou au lien entre certains traitements comme l'hydroxyurée et l'évolution de la maladie (Lottenberg et Hassell, 2005 ; Strouse *et al.,* 2008), ou enfin en fonction des groupes de population que l'on étudie selon qu'on s'intéresse plutôt à l'épidémiologie de la drépanocytose en Europe ou en Afrique du Nord par exemple (Mears *et al*., 1981 ; de Montalembert, Guilloud-Bataille, Feingold et Girot, 1993).

4.3.3. L'explication évolutionnaire

En première intention, on peut caractériser la médecine évolutionnaire comme l'application de la théorie de l'évolution en médecine. L'idée que la théorie de l'évolution peut avoir un impact sur les concepts de santé et de maladie n'est pas neuve (Bynum, 1983, 2002 ; Zampieri, 2009 ; Méthot, 2009), mais ce sont les travaux

du biologiste George Williams et du psychiatre Randolf Nesse (Williams et Nesse, 1991 ; Nesse, 1994, 2001, 2008) qui ont présenté la médecine évolutionnaire comme une « révolution » dans notre compréhension de la vulnérabilité des êtres humains aux maladies, au point de pasticher la célèbre affirmation de Dobzhansky (Dobzhansky, 1973) et d'affirmer que « rien ne fait sens en médecine si ce n'est à la lumière de la théorie de l'évolution » (Nesse, 1994). La théorie de l'évolution pourrait ainsi être appliquée à toutes les maladies[34], aurait vocation à devenir « le fondement scientifique » de la médecine (Nesse, 2008 ; Nesse *et al.*, 2010) et à occuper une place déterminante dans le curriculum de tout médecin. Reprenant des distinctions d'Ernst Mayr (Mayr, 1961), Nesse et William définissent donc la médecine évolutionnaire comme la recherche des causes lointaines et historiques de la maladie, répondant à la question « comment ? » (« *how ?* ») par opposition à la recherche des causes prochaines et fonctionnelles qui font l'objet de l'explication physiopathologique et répondent à la question « pourquoi ? », au sens de « comment les choses en sont arrivées là ? » (« *how come ?* »).

La médecine évolutionnaire doit d'abord expliquer l'existence, la naissance et la persistance de certaines maladies humaines, en se fondant sur le programme adaptationniste qui considère que les fonctions et la forme de l'organisme humain sont le résultat d'adaptations à des environnements qui ont varié très rapidement depuis le Pléistocène, en partie parce que l'homme lui-même a un impact considérable sur son environnement (Nesse et Stearns, 2008). Par rapport à ces variations rapides de l'environnement, la sélection naturelle agit lentement, ce qui explique l'existence d'un décalage (*mismatch*) entre nos gènes et l'environnement qui entraîne une série de « maladaptations » : ainsi, le mécanisme adaptatif de l'anxiété, supposé entraîner une réaction de fuite devant un prédateur, se déclencherait sans cause identifiée dans les troubles anxieux. Par ailleurs, la sélection naturelle est limitée par un certain nombre de contraintes évolutives, ce qui explique que nos organismes ne soient pas parfaitement optimisés mais représentent un ensemble de compromis (*trade-off*) entre adaptation et contrainte évolutive. Ainsi, les maladies auto-immunes sont en

[34] Il suffit de regarder le sommaire de l'ouvrage *Evolutionary Medicine and* Health (Trevathan *et al. éds.,* 2008), qui s'intéresse aussi bien aux maladies chroniques (obésité), qu'aux maladies ou aux mécanismes gynécologiques, aux maladies cardiovasculaires, etc.

quelque sorte le prix à payer pour avoir un organisme capable de se défendre par un système immunitaire. Enfin, la sélection naturelle a pour objectif de maximiser la capacité de reproduction des individus et non leur santé : c'est pourquoi certaines maladies qui apparaissent en moyenne après l'âge de la reproduction persistent, comme le cancer du sein, par exemple. Il est crucial de souligner que dans cette perspective, l'explication évolutionnaire aurait aussi un intérêt clinique : elle permettrait ainsi de distinguer entre les symptômes des maladies qu'il est souhaitable de traiter et ceux qui peuvent être compris comme des mécanismes adaptatifs de défense qu'il faudrait au contraire respecter (comme la fièvre).

Il est certain que l'explication évolutionnaire apparaît assez satisfaisante, pour des maladies comme la drépanocytose. Ainsi, l'explication évolutionnaire permet de comprendre que la drépanocytose se soit maintenue avec une prévalence élevée dans les populations subsahariennes, alors que les individus atteints de la maladie ont un taux élevé de mortalité et un faible taux de reproduction. Cette persistance s'explique comme un exemple de « fitness de l'hétérozygote » : si le fait d'être homozygote est très défavorable pour les individus, l'hétérozygotie pour la drépanocytose confère un avantage sélectif par rapport au paludisme. En effet, les individus hétérozygotes pour la drépanocytose seront moins sensibles au paludisme, maladie également endémique et mortelle de l'Afrique subsaharienne, du fait de la déformation des globules rouges qui empêche la pénétration du parasite dans les érythrocytes (Cavalli-Sforza et Bodmer, 1971). Il est à noter d'ailleurs que pour l'explication évolutionnaire, comme pour l'explication épidémiologique, nous sommes face à une explication populationnelle, qui utilise d'autres groupes de population comme classes de contraste avec la population étudiée. Dans l'explication épidémiologique, ces populations de contraste peuvent être soit d'autres populations où la distribution de la maladie est différente (comme les pays européens par exemple), soit des individus sains dans la population d'intérêt (quelle est la distribution des variants alléliques pour le gène codant l'hémoglobine dans les populations d'Afrique subsaharienne ne présentant pas la maladie ?). Dans l'explication évolutionnaire de la drépanocytose, c'est bien en comparant la distribution allélique théorique à la distribution allélique observée dans la population qu'on obtient le point de départ du raisonnement.

Cependant, comme le fait très justement remarquer Pierre-Olivier Méthot (Méthot, 2009), la médecine évolutionnaire telle qu'elle est présentée par Nesse et Williams, et qu'on appelle « médecine darwinienne », rencontre plusieurs difficultés conceptuelles[35]. Mais un autre style de médecine évolutionnaire existe, que Méthot appelle « médecine évolutive ». Contrairement à la médecine darwinienne qui propose des explications rétrogrades qui cherchent à expliquer les maladies contemporaines en prenant pour population de référence les populations humaines du Pléistocène, la médecine évolutive emploie des explications antérogrades destinées à « prédire les effets de processus évolutionnaires actuels sur la santé et la maladie humaine dans les environnements contemporains (par exemple, les hôpitaux) » (Méthot, 2010, p. 75) et qui sera utilisée par exemple pour expliquer le développement des résistances bactériennes aux antibiotiques. La médecine évolutive est une façon moins radicale d'adopter une perspective évolutionnaire sur les maladies : par opposition à la médecine darwinienne qui considère que la médecine doit être repensée à la lumière de la théorie de l'évolution, la médecine évolutive utilise des concepts moins conflictuels. Par exemple, même si le concept d'adaptation reste utile pour étudier l'évolution des relations hôte-parasite, il n'y a pas d'adhésion à un programme adaptationniste visant à considérer que toutes les fonctions et les formes de l'organisme humain sont le résultat d'adaptation à un environnement changeant par le biais de la sélection naturelle. Dans cette perspective, *« l'évolution est un outil analytique permettant d'apporter un éclairage nouveau dans certains contextes théoriques et pratiques déterminés. »* (Méthot, 2009, p. 963)

Il faut noter que la médecine évolutive a également des conséquences thérapeutiques : comprendre que l'essor des infections nosocomiales est en partie due à l'utilisation massive d'antibiotiques qui favorise l'émergence de résistances bactériennes aux antibiotiques du fait de la vitesse à laquelle la sélection naturelle agit sur les souches bactériennes qui ont un fort taux de reproduction et une durée de vie courte, a entraîné la mise en place de stratégies thérapeutiques différentes, consistant

[35] Les plus fréquentes sont : critique du programme adaptationniste proposée par Lewontin et Gould (Gould et Lewontin, 1979), critique du choix du Pléistocène comme période de référence pour évaluer les changements de l'espèce humaine en termes de sélection naturelle, etc. Outre ces critiques, on peut également se demander si des explications évolutionnaires aussi satisfaisantes que celle de la drépanocytose sont susceptibles d'exister pour toutes les maladies

par exemple à limiter l'usage des antibiotiques quand ceux-ci ne sont pas strictement nécessaires, ou à permuter le type d'antibiotiques utilisé.

Ainsi, si les prétentions de la médecine darwinienne à devenir un fondement théorique et scientifique pour la médecine peuvent paraître démesurées, nul doute qu'il existe une variété d'explications évolutionnaires des maladies et que celles-ci se caractérisent *a minima* comme des explications populationnelles de la maladie qui font appel à la théorie de l'évolution pour comprendre les causes historiques et évolutives des maladies, éclairent la compréhension de certains symptômes compris comme des mécanismes adaptatifs de défense et ont des conséquences thérapeutiques.

4.3.4. L'explication physiopathologique

La question « pourquoi cette maladie advient-elle ? » peut être comprise au sens évolutionnaire que nous avons explicité plus haut, mais elle peut aussi être comprise au sens fonctionnel (« comment la maladie advient-elle ? ») : elle appelle alors une explication physiopathologique, fondée sur la mise en évidence d'une série de mécanismes. Machamer, Darden et Craver définissent les mécanismes comme *« des entités et des activités organisées de telle sorte qu'elles produisent des changements réguliers à partir de conditions initiales ou premières jusqu'à des conditions terminales ou finales »* (Machamer, Darden et Craver, 2000, p. 3). Une telle définition met l'accent sur la nécessité dans l'explication mécaniste de décomposer un phénomène en des entités et des activités. Par exemple dans le cas de la drépanocytose, les entités sont l'ensemble des entités impliquées dans la transcription et la traduction de l'allèle HbS en protéine de l'hémoglobine mutée, mais aussi les globules rouges déformés et les vaisseaux dans lesquels les globules rouges déformés vont circuler, et ainsi de suite. Les activités désignent par exemple l'activité de traduction, l'activité de transcription, le changement de conformation de la protéine, le processus de polymérisation de l'hémoglobine qui entraîne la déformation des hématies, etc. Cette définition met aussi l'accent sur la nécessité dans l'explication mécaniste de localiser dans le temps (conditions initiales et conditions finales) et dans l'espace le phénomène en question.

Nous voudrions ici insister sur la multiplicité des mécanismes qui peuvent être impliqués dans une explication physiopathologique. En effet, dans les manuels de médecine, on présente souvent un raccourci de l'explication physiopathologique de la drépanocytose, en se focalisant sur la mutation homozygote de l'allèle HbS, qui entraîne une conformation anormale de la chaîne beta de l'hémoglobine, laquelle résulte en une capacité de polymérisation de l'hémoglobine mutée qui va altérer la forme normale des hématies et leur donner un aspect en faucille. Cette description mécaniste est juste mais a) elle peut elle-même être décomposée en une série de mécanismes : par exemple, on pourrait détailler les mécanismes moléculaires par lesquels l'allèle HbS muté va être transcrit puis traduit en protéine de l'hémoglobine mutée b) cette explication physiopathologique ne suffit pas à expliquer l'ensemble des symptômes de la drépanocytose et l'ensemble des complications aiguës ou chroniques qui peuvent en résulter. Ainsi, une fois qu'on a expliqué la déformation des hématies des patients homozygotes pour la drépanocytose, il faut encore mettre à jour les mécanismes qui relient la déformation de l'hémoglobine aux trois grands symptômes de la drépanocytose (l'anémie hémolytique, les crises vaso-occlusives, la susceptibilité aux infections), les symptômes chroniques de la drépanocytose (retard de croissance et retard pondéral, carences nutritionnelles en folates, retard pubertaire fréquent, troubles cardiopulmonaires, splénomégalie, atteintes oculaires), et enfin les mécanismes qui expliquent l'évolution de la maladie dans le temps (apparition de la maladie quelques mois après la naissance après disparition de l'hémoglobine fœtale, facteurs favorisant les symptômes aigus, pronostic à long terme, complications aiguës ou chroniques de la maladie). Une explication mécaniste de la maladie est donc susceptible de toucher les quatre cibles explicatives (causes, symptômes, évolution et traitement).

Comme l'explication épidémiologique et l'explication évolutionnaire, l'explication mécaniste de la maladie fait également référence à des classes de contraste, que ce soit en termes de causes, de symptômes, d'évolution ou de traitement des maladies. Ainsi, l'explication du comportement pathologique de l'hémoglobine mutée fait explicitement référence au comportement normal de l'hémoglobine et l'explication d'un certain type d'hémoglobine mutée fait référence à la fois au comportement normal de l'hémoglobine et au comportement pathologique

des autres hémoglobines mutés. Ainsi des symptômes différents ou une évolution différente de la maladie sont expliqués en fonction de variations d'un même mécanisme, qui peut être soit tout à fait inopérant résultant en un nouveau mécanisme (la mutation de l'allèle HbS entraîne un changement de conformation de l'hémoglobine mutée qui empêche la formation normale de l'hématie et résulte en un nouveau mécanisme de formation de l'hématie en faucille par polymérisation), ou qui est partiellement inopérant (dans le cas de l'hétérozygote drépanocytaire par exemple), etc.

4.3.5. L'explication clinique, lieu d'instanciation des différents types d'explications

La question « pourquoi cet individu est-il malade ? » appelle deux types d'explications. La première est le diagnostic qui subsume les symptômes de la maladie d'un individu sous une entité qui a valeur d'explication. La deuxième, et c'est sur celle-ci que nous nous concentrerons, est une explication clinique : il s'agit de comprendre pourquoi ce malade particulier est atteint de cette maladie. C'est donc un cas d'instanciation : en fonction de l'anamnèse et de l'histoire familiale, personnelle et médicale récente d'un patient, on va choisir dans l'ensemble des causes susceptibles d'expliquer la maladie-type quelles causes expliquent le cas du patient. Dans le cas de la drépanocytose, le processus d'instanciation est relativement simple pour la maladie en général : il suffit de caractériser le statut du patient (homozygote, hétérozygote, hétérozygotie composite). Mais pour comprendre la symptomatologie particulière d'un patient drépanocytaire qui fait une crise vaso-occlusive, le clinicien cherchera à instancier un ensemble de causes déclenchantes possibles : présence d'une infection sous-jacente, situation d'hypoxie relative, stress, froid, etc. Nous ne reprendrons pas en détail la caractérisation de l'explication individuelle de la maladie que nous avons déjà décrite plus haut à travers le processus d'instanciation. Nous nous contenterons seulement de faire trois remarques pour compléter cette caractérisation.

La première remarque concerne le fait que l'instanciation suppose de faire appel à des classes de contraste : on explique les particularités de la drépanocytose de Julia par référence aux autres formes de drépanocytose (autres formes de syndrome drépanocytaire majeur, etc.), mais aussi par référence aux autres individus singuliers

atteints de drépanocytose. On ne pourrait d'ailleurs pas avoir d'explication de la drépanocytose de Julia sans la connaissance d'autres cas de drépanocytose qui permettent de comprendre quelles sont les différences et quelles sont les particularités de la drépanocytose de Julia. La deuxième remarque concerne le degré de généralité de l'explication clinique : contrairement aux explications évolutionnaire, épidémiologique et physiopathologique, l'explication clinique est l'explication d'un fait particulier (la maladie individuelle en tant qu'instance) et non pas d'une classe de faits (la maladie individuelle en tant que maladie-type, une classe de maladie, la maladie en général).

Finalement, la troisième remarque concerne le caractère spontanément intégratif de l'explication clinique. Nous avons montré que chaque type d'explication précédemment présenté pouvait intervenir à la fois dans l'explication des causes, des symptômes, de l'évolution et du traitement de la maladie-type, mais qu'à chaque explication correspondait une forme particulière de causalité et que chaque explication apportait des informations particulières quant aux différentes cibles explicatives de la maladie-type. De façon remarquable, ces trois formes d'explication sont convoquées à des degrés et à des moments divers dans l'explication clinique : (a) au moment de l'observation clinique, à travers le recueil des caractéristiques personnelles (âge, sexe, ethnie, antécédents médicaux) ou familiales du patient (antécédents familiaux, type de mutation retrouvée dans la famille), ou à travers l'histoire de la maladie (âge d'apparition des premiers symptômes, type et fréquence des symptômes, facteurs déclenchants des complications aiguës, type de traitement utilisé) (b) au moment des différents tests diagnostics (par exemple, étude de l'hémoglobine permettant de caractériser qualitativement et quantitativement les différents types d'hémoglobine du patient) (c) au moment de l'instauration d'un traitement (utilisation des protocoles de traitement adaptés à la caractérisation de la forme de drépanocytose présentée par le patient et à l'efficacité des traitements précédemment utilisés) et enfin (d) dans le suivi du patient tout au long de sa vie dans le cas des maladies chroniques comme la drépanocytose.

4.3.6. Il n'y a pas de hiérarchie entre les différents types d'explication en médecine

C'est précisément parce que ces différents types d'explication ont vocation à s'appliquer à toutes les maladies (en dépit des quelques réserves que nous avons discutées pour les explications évolutionnaires) et parce que l'explication clinique permet d'intégrer spontanément les trois types d'explications (épidémiologique, évolutionnaire, physiopathologique) que l'on comprend qu'il n'y a pas de sens à admettre une hiérarchie entre ces types d'explications. Il n'y a pas de sens à dire qu'une explication physiopathologique de la drépanocytose est plus importante qu'une explication épidémiologique, puisque chacune de ces explications est utile à l'explication clinique d'une maladie-instance. D'ailleurs, chacune de ces explications intervient dans plusieurs cibles explicatives d'une maladie-type (causes, symptômes, évolution, traitement) même si elles participent différemment à l'explication de ces cibles explicatives. C'est ainsi que pour les causes de la maladie, une explication épidémiologique mettra en évidence des facteurs de risque individuels à la maladie, une explication évolutionnaire mettra en évidence des causes évolutives, une explication physiopathologique soulignera la causalité mécanique. Ces explications sont également fortement dépendantes les unes des autres (Lemoine, 2011, 185–211) : par exemple, l'explication évolutionnaire de la drépanocytose s'appuie sur la description épidémiologique de la maladie (prévalence élevée en région subsaharienne, région endémique de la drépanocytose et du paludisme) et sur l'explication physiopathologique (maladie déclenchée par une anomalie moléculaire sur le gène HbS). De la même manière, pour formuler une explication physiopathologique, il faut être confronté à plusieurs cas de la même maladie : l'explication physiopathologique nécessite donc une certaine forme d'explication épidémiologique pour avoir un degré de généralité satisfaisant.

Pour reprendre les termes de Maël Lemoine, on peut donc dire que chacune de ces explications est en un sens complète par rapport au type de question à laquelle elle répond mais complémentaire par rapport aux autres types d'explication – ce qui est bien conforme au pluralisme explicatif des sciences médicales que nous défendons. A chaque type de question correspond un type d'explication qui se distingue par la

méthode qu'elle déploie (méthodes statistiques, explications mécanistes, etc.), et par le type de classes de contraste par rapport auxquelles elle se constitue, et qui peut être plus ou moins satisfaisante, ou plus ou moins empiriquement démontrée mais dont la forme suffit à répondre à la question posée.

4.4. Jusqu'à quel point une théorie médicale doit-elle expliquer ?

Nous avons à présent défini les objets possibles d'une explication médicale et les différents types d'explication possibles dans les sciences médicales. Il s'agit à présent de déterminer jusqu'à quel point une théorie médicale doit expliquer, c'est-à-dire de décrire quelles sont les conditions *a minima* et *a maxima* d'une théorie médicale. Nous introduirons également dans cette section le concept de théorie médicale partielle, un concept fondamental pour comprendre ce que pourrait être une théorie génétique de la maladie.

4.4.1. Définition d'une théorie médicale a minima *et avantages de notre conception*

Nous avons identifié plusieurs objets d'une explication (la maladie individuelle en tant que maladie-type, la maladie individuelle en tant que maladie-instance, une classe de maladie ou la maladie en général), quatre cibles explicatives d'une théorie médicale (les causes, les symptômes, l'évolution et le traitement d'une maladie) et quatre formes d'explication (explication épidémiologique, explication évolutionnaire, explication physiopathologique et explication individuelle) dont nous avons montré qu'elles pouvaient intervenir dans l'explication de différentes cibles explicatives. Toute caractérisation d'une théorie médicale peut donc se faire en fonction de ces différentes dimensions : objet, niveau, cible explicative, type d'explication (voir Figure 18).

4 critères pour caractériser une théorie médicale	
Quel est l'objet de la théorie ?	La maladie ? Une classe de maladie ? Une maladie ?
À quel niveau la théorie explique-t-elle ?	Théorie générale ? Théorie régionale ? Théorie d'une maladie ?
Quelles sont les cibles explicatives de la théorie ?	Causes Symptômes Évolution Traitement
Quel type d'explication la théorie mobilise-t-elle ?	Explication épidémiologique Explication évolutionnaire Explication mécaniste ou physiopathologique

Figure 18 : Quatre critères pour caractériser une théorie médicale. Notre concept de théorie médicale suppose de se poser quatre questions : quel est l'objet de la théorie ? À quel niveau la théorie explique-t-elle ? Quelles sont les cibles explicatives de la théorie ? Quel type d'explication la théorie mobilise-t-elle ?

Si nous reprenons maintenant la première condition (définie en 4.1.) d'identification d'une théorie qui est l'explication d'une ou de plusieurs classes de faits (et non d'un fait individuel), et qu'on admet qu'il n'y a pas de hiérarchie entre les différents objets-maladie ou entre les différents types d'explications de la maladie, on aboutit à la définition suivante d'une théorie médicale *a minima* : *une théorie médicale a minima suppose au minimum un type d'explication (évolutionnaire, épidémiologique, ou pathophysiologique) d'au minimum une cible explicative (causes, symptômes, évolution, ou traitement) d'une maladie individuelle en tant que maladie-type, ou d'une classe de maladie ou de la maladie en général.*

Cette définition permet de comprendre que l'explication clinique d'une maladie individuelle en tant que maladie-instance est donc la seule forme d'explication qui ne puisse se qualifier en tant que théorie médicale – précisément parce que l'explication

clinique est une explication d'un fait particulier, ponctuel et singulier (la maladie d'un individu), par opposition aux trois autres formes d'explications qui concernent toutes des classes de faits indépendants.

Cette façon de définir le concept de théorie médicale *a minima* présente plusieurs avantages, qui répondent aux insuffisances que nous avons identifiées dans les conceptions de Schaffner, Thompson, Kazem Sadegh-Zadeh et Thagard et aux exigences que nous avons formulées au début de ce chapitre.

(1) Tout d'abord, c'est une conception qui répond à notre démarche minimalement normative en établissant des critères nécessaires à l'identification d'une théorie médicale mais relativement peu exigeants et qui sont susceptibles d'être combinés de différentes façons (une théorie d'une maladie-type mobilisant plusieurs formes d'explications de la même cible explicative, une théorie d'une classe de maladie mobilisant une seule forme d'explication pour des cibles explicatives différentes, etc.), aboutissant à un spectre des formes possibles de théorie médicale.

(2) Ce faisant, cette définition d'une théorie médicale permet de refléter assez fidèlement l'usage historique du terme de théorie médicale, mais elle est surtout paradoxalement compatible avec des versions modérées du scepticisme vis-à-vis des théories médicales, en particulier avec ceux qui considèrent, comme Harold Kincaid que la médecine produit des « petits bouts d'explication causale[36] » (Kincaid, 2008) Ainsi, on peut caractériser par exemple la théorie des germes comme une théorie régionale d'une classe de maladie, employant une explication physiopathologique pour expliquer les causes des maladies infectieuses, ou la théorie évolutionnaire de la drépanocytose comme une théorie d'une maladie-type, employant une explication évolutionnaire pour caractériser les causes évolutives de la drépanocytose, etc.

(3) Cette conception des théories médicales conserve les propriétés générales des théories biologiques, que nous empruntons à Schaffner dans la section 4.2.1 (théories de moyenne portée, théories orientées vers l'objet du vivant, théories

[36] Nous avons soutenu plus haut que notre conception des théories médicales n'était pas réductible à une explication causale des maladies. Ceci n'est pas contradictoire avec l'affirmation de Kincaid qui cherche moins à défendre le caractère intrinsèquement causal des théories médicales qu'à discuter la capacité ou la nécessité pour les explications médicales d'unifier de larges classes de faits (se reporter à l'annexe 2 pour une discussion complémentaire).

fondamentalement interdisciplinaires) tout en respectant la spécificité de l'objet des théories médicales qui est l'objet maladie.

(4) Cette conception des théories médicales permet de souligner le caractère fondamentalement interdisciplinaire des explications médicales, tout en distinguant ce qui relève de modèles empruntés à d'autres sciences utilisés dans une explication médicale et qui sont des théories *en médecine* (comme par exemple les lois de Mendel dans l'explication des maladies génétiques ou le modèle insuline glucose de Bolie dans une explication générale du diabète) et ce qui relève à proprement parler de théories médicales.

(5) Elle est tout à fait compatible avec le pluralisme explicatif dans les sciences médicales qu'elle conforte en identifiant plusieurs types de questions et plusieurs formes d'explications qui peuvent correspondre aux intérêts des différents acteurs impliqués dans la médecine (chercheurs, praticiens, etc.).

(6) Elle a l'avantage de permettre une approche pluraliste du concept de causes de la maladie, en distinguant les facteurs de risques, les causes évolutionnaires, les causes mécanistes et l'instanciation d'un réseau causal.

(7) Enfin, cette conception des théories médicales permet de souligner qu'il n'existe pas que des explications causales de la maladie – les causes de la maladie n'étant qu'une cible explicative parmi d'autres. Ainsi sont mises en valeur d'autres formes d'explications qui admettent un certain niveau de généralité et qui intéressent de très près le clinicien, mais que les philosophes ont souvent délaissées.

4.4.2. Trois classes de faits supplémentaires pour évaluer la précision d'une théorie médicale

Si la définition d'une théorie médicale *a minima* peut se déduire assez naturellement, en combinant la notion d'explication de classes de faits indépendants et la description des différentes dimensions des explications médicales, il est plus difficile de définir une théorie *a maxima,* c'est-à-dire une forme de théorie « idéale ».

En effet, une première réponse pourrait être de définir la théorie médicale *a maxima* comme la théorie avec le plus haut niveau de généralité (une théorie de la maladie en général), susceptible d'expliquer toutes les cibles explicatives de la maladie en général et en combinant de façon complémentaire les explications

épidémiologique, évolutionnaire et physiopathologique. Une telle définition n'a cependant pas de sens pour la simple raison qu'il y a un compromis nécessaire à trouver entre l'objet et le niveau d'une théorie d'une part et la précision de l'explication qui peut être fournie d'autre part. Par exemple, que ce soit dans la description des causes évolutionnaires, épidémiologiques ou mécanistes, on ne peut attendre le même niveau de précision d'une théorie d'une classe de maladie et d'une théorie d'une maladie-type. La théorie des germes permet d'identifier une même cause pour l'ensemble d'une classe de maladie, mais elle ne peut pas décrire dans le même temps les mécanismes qui interviennent dans cette classe de maladie, avec la même précision qu'on attendra en revanche d'une explication de la tuberculose en tant que maladie-type.

Il nous faut donc trouver des classes de faits supplémentaires, qui soient en lien avec les quatre cibles explicatives principales d'une maladie (les causes, les symptômes, l'évolution et le traitement), qui soient susceptibles d'être la cible d'explications épidémiologique, évolutionnaire ou physiopathologique, et dont l'explication permette d'augmenter la précision d'une théorie et de comparer deux théories de même niveau entre elles. Nous identifions trois classes de faits supplémentaires qui répondent à ces critères : (1) la variabilité interindividuelle, (2) la variabilité intra-individuelle (3) la comorbidité.

4.4.2.1. La variabilité interindividuelle

La notion de variabilité interindividuelle recouvre l'ensemble des variations de l'expression d'une maladie, que ce soit du point de vue des causes de la maladie, des symptômes, de l'évolution ou des traitements. Un exemple d'explication de la variabilité interindividuelle est le problème des porteurs sains ou des individus asymptomatiques – c'est-à-dire des individus qui semblent exposés aux mêmes causes de la maladie que les individus malades mais qui ne développent pas la maladie pour autant. L'expression de « porteurs asymptomatiques » est fréquemment utilisée en infectiologie, depuis le célèbre exemple de Marie Typhoïde, identifiée aux États-Unis comme porteuse saine de la fièvre typhoïde et rendue responsable de 51 cas de contamination avérée et de trois décès. On retrouve aussi cette notion en génétique

dans le cas des maladies mendélienne à transmission récessive, pour qualifier les porteurs hétérozygotes asymptomatiques. Mais on peut en fait élargir cette notion à de nombreuses classes de maladie (les fumeurs qui ne développent pas le cancer du poumon, les porteurs de l'antigène HLA B57 qui ne développent pas la spondylarthrite ankylosante, etc.).

Mais la variabilité interindividuelle désigne aussi la variabilité phénotypique dans l'expression des symptômes d'une maladie et de son évolution dans le temps. Ainsi, la définition des profils phénotypiques des drépanocytaires est un enjeu majeur de la recherche actuelle sur la drépanocytose. Si un certain nombre de mécanismes ont été établis qui permettent de comprendre les liens entre l'anomalie moléculaire et les différents types de symptômes de la drépanocytose, on dispose de très peu de facteurs prédictifs ou de mécanismes pour comprendre pourquoi certains patients ne manifesteront qu'une partie de ces symptômes au cours de la vie, pourquoi pour certains la fréquence des symptômes diminue avec l'âge alors que pour d'autres la fréquence des symptômes augmente ou se stabilise. Encore une fois, ce problème de la variabilité de l'expression interindividuelle se pose pour de nombreuses autres maladies : une des difficultés majeures dans la recherche sur le cancer est ainsi de comprendre pourquoi certains cancers évoluent vers le stade métastatique ou les déterminants pronostics des rechutes.

Enfin la variabilité interindividuelle concerne de toute évidence la sensibilité aux traitements : il s'agit alors d'expliquer pourquoi des patients atteints de la même maladie répondent de manière radicalement différente aux mêmes traitements.

4.4.2.2. La variabilité intra-individuelle

La variabilité intra-individuelle peut paraître un problème plus secondaire mais son explication est pourtant tout à fait décisive pour le clinicien et elle est particulièrement évidente dans les maladies chroniques. Il s'agit cette fois-ci de comprendre pourquoi le même individu ne manifeste pas les mêmes symptômes dans des circonstances similaires. Par exemple, comment comprendre qu'un même patient drépanocytaire ne va pas réagir de la même manière à des environnements similaires supposés déclencher une crise vaso-occlusive ? Ou comment comprendre le

développement de résistances à certaines lignes de traitement dans des maladies chroniques, là où le traitement a fait ses preuves quelques mois plus tôt ?

4.4.2.3. La comorbidité

Enfin, la comorbidité est un terme ambigu qui désigne en première intention le fait que deux ou plusieurs maladies apparaissent simultanément chez un même individu. La question décisive qui se pose est alors de savoir à quel point cette comorbidité est une coïncidence ou révèle un lien (facteur d'exposition, mécanisme co-évolutif, mécanisme physiopathologique) commun entre les deux maladies. Par extension, la comorbidité désigne donc la probabilité qu'un malade atteint d'une maladie A soit atteint d'une maladie B à un moment donné de son existence. La question de la comorbidité est tout à fait passionnante parce qu'elle a des conséquences aussi bien en termes de prévention et de santé publique, qu'en termes de traitements proposés. Elle éclaire par ailleurs un certain nombre d'affirmations répandues dans le milieu médical selon lesquelles on ne soigne pas des maladies, mais des patients. Cette affirmation peut être comprise comme la négation de l'existence des maladies, mais elle peut aussi être comprise dans un sens beaucoup moins radical comme l'expression de l'activité clinique où les médecins ne sont pas confrontés à « la forme typique » d'un cas exemplaire décrit dans un manuel de médecine, mais à un patient qui présente un certain nombre de particularités, dont des antécédents médicaux et des maladies actives en cours de traitement avec lesquelles il faut savoir composer lorsqu'on instaure un nouveau traitement pour une nouvelle maladie.

4.4.2.4. Trois arguments en faveur de ces trois classes de faits

Si nous avons identifié ces trois classes de faits comme des classes de faits susceptibles d'augmenter la précision d'une théorie médicale, c'est pour trois raisons. (1) Tout d'abord, nous avons établi dans la section précédente que les explications épidémiologique, évolutionnaire et physiopathologique sont trois formes d'explication contrastive. Or, expliquer chacune de ces trois classes de faits suppose – et ce, quel que soit le type d'explication choisi – un raffinement des classes de contrastes. Plus on est capable de distinguer différentes classes de contraste dans l'explication d'un même

phénomène et plus on est proche des causes ou des mécanismes responsables des différences observées. Ainsi, expliquer la variabilité interindividuelle des symptômes de la drépanocytose signifie qu'on est capable d'expliquer non seulement pourquoi les individus atteints de la drépanocytose sont malades, mais aussi pourquoi certains individus hétérozygotes présentent certains symptômes de la drépanocytose tandis que d'autres individus hétérozygotes n'en présentent pas, et pourquoi certains patients vont présenter des symptômes et une forme d'évolution plus sévères que d'autres, alors même qu'ils présentent les mêmes mutations du gène de l'hémoglobine. De la même manière, expliquer la variabilité intra-individuelle d'un patient à la drépanocytose permet d'expliquer non seulement l'ensemble de ces symptômes à un moment donné, mais aussi leur diversité et leur évolution. Enfin, déterminer des profils de comorbidité pour différentes classes de patients atteints de la drépanocytose permet également de raffiner la compréhension des causes, de l'évolution, des symptômes et du traitement de la maladie. Ces trois concepts de variabilité inter-individuelle, variabilité intra-individuelle et comorbidité sont donc intrinsèquement liés à l'identification de nouvelles classes de contraste dans les explications épidémiologique, évolutionnaire ou physiopathologique d'une maladie donnée.

(2) Par ailleurs, ces trois classes de faits interviennent dans l'explication de toutes les cibles explicatives de la maladie. On peut ainsi parler de variabilité inter-individuelle aussi bien pour les causes d'une maladie (tous les patients atteints d'une même maladie ne le sont pas pour les mêmes causes comme le montrait l'exemple de l'ulcère), que pour les symptômes (tous les patients atteints d'une même maladie ne présentent pas les mêmes symptômes), l'évolution (tous les patients atteints d'une même maladie ne présentent pas la même évolution de leur maladie) ou le traitement (tous les traitements ne sont pas efficaces de la même manière pour tous les patients atteints d'une même maladie).

(3) Enfin ces trois classes de faits peuvent être décrites, quel que soit l'objet-maladie que l'on cherche à expliquer et quel que soit le niveau de la théorie qu'on examine. Ainsi, on peut aussi bien examiner comment une théorie de la tuberculose rend compte de la variabilité interindividuelle, de la variabilité intra-individuelle et de

la comorbidité de la tuberculose qu'examiner ces trois mêmes critères pour une théorie des maladies infectieuses.

Ces trois classes de faits permettent donc d'abord d'évaluer le degré de précision d'une théorie donnée et ce, quelle que soit la forme que prend cette théorie médicale, que ce soit une théorie d'une maladie-type qui explique avec un seul type d'explication une cible explicative de la maladie donnée ou que ce soit une théorie d'une classe de maladie qui intègre plusieurs types d'explications de plusieurs cibles explicatives de la classe de maladie en question, etc. Mais ces trois classes de faits permettent du même coup de comparer entre elles deux théories de même niveau, qui sont en compétition dans l'explication d'un même objet-maladie.

Ainsi, si la définition de la théorie médicale *a minima* découle assez naturellement des différents critères d'identification d'une théorie scientifique et des dimensions d'une explication médicale que nous avons exposés précédemment, il n'en va pas de même de la définition d'une théorie médicale *a maxima*, le niveau de précision d'une théorie étant en partie sacrifié à mesure que l'objet de la théorie est plus général. En revanche, la capacité à expliquer les trois classes de faits que sont la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité, sont des facteurs de précision qui permettent à la fois d'évaluer le niveau de précision d'une théorie donnée et de comparer des théories de même niveau entre elles.

4.4.3. La notion de théorie partielle

Nous voudrions maintenant introduire un dernier élément de terminologie : le concept de « théorie partielle ». Il y a bien sûr un sens trivial dans lequel toute théorie peut être dite partielle, au sens où toute théorie suppose un travail d'abstraction et d'idéalisation qui néglige certains détails qui ne sont pas pertinents en fonction du type de question posée ou en fonction de ce que l'on souhaite faire de la réponse.

Dans un second sens, moins trivial, le caractère partiel des théories médicales est intrinsèquement lié à la pluralité des explications que nous avons reconnues être à l'œuvre dans les sciences médicales et qui, si elles sont complètes par rapport à l'objet de leur explication, sont néanmoins partielles au regard d'une compréhension aussi extensive qu'on peut l'imaginer de la maladie. En ce sens, le fait même de définir une

théorie médicale *a minima* est une reconnaissance du caractère partiel des théories médicales.

Mais c'est encore dans un troisième sens que nous posons le concept de théorie partielle, un sens qui doit être compris à la fois dans le cadre de l'interactionnisme co-constructionniste que nous avons posé comme horizon général de l'explication des maladies et dans le cadre analytique d'une stratégie de décomposition des phénomènes. Les maladies sont des phénomènes complexes à expliquer au sens où ils impliquent des interactions permanentes entre des causes de différente nature (déterminants sociaux, environnementaux, génétiques, épigénétiques de la maladie) et qui interviennent à différents niveaux (moléculaires, cellulaires, tissulaires, phénotypiques) et à différents stades (développement, croissance, vieillissement) de la vie d'un organisme. Ce point étant admis, nous faisons l'hypothèse qu'analyser le phénomène du point de vue d'une cause particulière peut permettre de progresser dans la compréhension du phénomène, même si cette explication n'est que partielle (elle ne constitue pas une explication exhaustive du phénomène), éventuellement temporaire (dans l'attente d'une explication qui rende compte des interactions entre cette cause et les autres déterminants du phénomène) et limitée (au sens où elle ne peut expliquer qu'une partie du phénomène). C'est ainsi que l'hypothèse d'une théorie génétique que nous avons faite peut être comprise comme un exemple de théorie partielle et c'est ce que nous allons à présent détailler.

4.5. A quoi ressemblerait une théorie génétique ?

Nous avons fait l'hypothèse que la généticisation pouvait correspondre à l'élaboration d'une théorie génétique. Pour justifier la formulation de notre hypothèse, nous avons d'abord examiné les objections qui sont opposées le plus fréquemment au concept de théorie médicale avant de proposer un ensemble de critères qui permettent de caractériser les différentes formes de théorie médicale qui existent, une définition *a minima* d'une théorie médicale, trois critères permettant l'évaluation du degré de précision d'une théorie médicale et un concept de théorie médicale partielle.

Nous pouvons à présent revenir à notre hypothèse d'une théorie génétique et appliquer cet arsenal conceptuel pour déterminer à quoi ressemblerait une théorie génétique, si tant est qu'elle existe. Nous procéderons en trois temps. Nous justifierons d'abord l'intérêt d'une théorie génétique, non pas seulement comme nous l'avons déjà fait, au motif que le concept de maladie génétique a disparu et que la généticisation ne peut être réduite au génocentrisme, mais aussi parce que le gène est un composant récurrent dans différents types d'explication des maladies. Dans un deuxième temps, nous présenterons les critères pour identifier une théorie génétique *a minima* et pour évaluer son degré de précision. Enfin, nous présenterons une typologie non exhaustive mais suffisamment représentative des différents types de théorie génétique qu'il serait possible d'imaginer.

4.5.1. Les gènes comme explanans *dans les explications évolutionnaire, épidémiologique et physiopathologique d'une maladie*

Nous avons déjà exposé un certain nombre de raisons pour lesquelles nous avons formulé l'hypothèse d'une théorie génétique. Il s'agit de rendre compte du double processus d'effondrement du concept de maladie génétique et d'extension de ce concept à toutes les maladies, c'est-à-dire de la généticisation des maladies. Mais à présent que nous avons caractérisé les différents types d'explication qui existent en médecine, le concept de théorie génétique présente un nouvel avantage : celui de permettre une intégration entre ces différents types d'explication des maladies du point du vue du gène, dans la mesure où le gène est impliqué dans les explications épidémiologique, évolutionnaire et physiopathologique de la maladie et que dans ces trois types d'explication, la nécessité d'explications interactionnistes est de plus en plus accentuée. Autrement dit, nous défendons l'idée qu'il peut exister une théorie partielle de la maladie centrée sur le gène en tant qu'il intervient dans les trois types d'explications médicales sus-identifiés, sans retour à une forme d'exceptionnalisme génétique.

En effet, le gène intervient d'abord dans les explications épidémiologiques où s'est d'ailleurs développée une sous-discipline de l'épidémiologie : l'épidémiologie génétique. Sans rentrer dans une analyse détaillée de l'émergence de l'épidémiologie

génétique[37] et si, à proprement parler, le terme d'épidémiologie génétique est apparu en 1978 (Morton et Chung, 1978), on peut caractériser l'épidémiologie génétique comme une convergence entre généticiens et épidémiologistes qui a commencé dans la seconde moitié du vingtième siècle (Neel et Schull, 1954) et dont la définition la plus célèbre est celle de Morton et Chung : *« une science qui traite de l'étiologie, de la distribution et du contrôle de la maladie dans des groupes d'apparentés, ainsi que des causes héritées de la maladie dans la population »* (Morton, 1978). Cette première définition semble intimement liée à la génétique mendélienne et à l'étude des facteurs génétiques de la maladie dans les populations humaines en tant que ceux-ci sont hérités. Toutefois, cette définition s'est modifiée au fur et à mesure de l'extension du concept de maladie génétique, de l'élargissement des techniques de génétique moléculaire et des méthodes d'analyse épidémiologique et de l'idée de plus en plus prégnante selon laquelle des déterminants génétiques sont impliqués dans toutes les maladies. C'est ainsi que Khoury propose une définition plus large de l'épidémiologie génétique :

> « Il est clair à partir de la discussion ci-dessus que l'épidémiologie génétique est la mieux caractérisée comme l'étude du rôle des facteurs génétiques et de leurs interactions avec les facteurs environnementaux dans l'apparition de la maladie dans les populations humaines. Les facteurs environnementaux sont définis largement comme les facteurs exogènes – chimiques, physiques, infectieux et nutritionnels. L'épidémiologie génétique est liée de près aux sciences biomédicales et au champ de la santé publique. Son rôle au sens large est de comprendre le rôle des facteurs génétiques dans l'étiologie de la maladie dans les populations humaines. Dans cette perspective, les types d'études disponibles (par exemple, les études familiales, les études de consanguinité ou les enquêtes de population) et les techniques statistiques (par exemple, l'analyse de ségrégation, l'analyse de liaison) doivent être considérées comme des moyens plutôt que comme une fin. » (Khoury *et al.,* 1993, p. 13)

[37] Pour une analyse de l'histoire de l'épidémiologie génétique, on pourra se reporter aux nombreux travaux de Khoury (Khoury, 1997 ; Khoury *et al.,* 1993, 2004) ou à ceux de Morton (Morton et Chung, 1978 ; Morton, 2006), dont la contribution a été décisive dans ce champ, ou encore à cet article qui analyse cinq articles scientifiques ayant constitué une étape importante de la constitution de l'épidémiologie génétique (Dewan, 2010).

Nous sommes donc dans une discipline qui, quelles que soient les discussions qui existent sur les limites ou l'étendue de ses méthodes (Smith *et al.,* 2005 ; Sharp, 2005 ; Ebrahim, 2006), analyse le rôle des variants génétiques impliqués dans les maladies humaines dans un cadre interactionniste (même si ce cadre est encore très proche de la génétique quantitative et donc d'un interactionnisme quantitatif).

Quant à l'explication évolutionnaire de la maladie, elle s'appuie (en particulier en ce qui concerne la médecine darwinienne) sur la théorie synthétique de l'évolution, telle qu'elle a émergé entre les années 1920 et les années 1950 et qui intègre à la théorie darwinienne la génétique mendélienne et la génétique des populations. Il est donc bien évident que les gènes jouent un rôle important dans l'explication évolutionnaire des maladies à un double titre. Ils sont impliqués dans l'explication des mécanismes évolutionnaires (qui incluent la sélection naturelle, mais aussi des phénomènes comme la dérive génétique aléatoire) et ils sont impliqués en tant qu'unité de sélection, c'est à dire en tant qu'unités sur lesquelles s'exerce la sélection naturelle. Or, dans cette perspective, il faut noter que là encore, aujourd'hui, c'est le cadre interactionniste qui prévaut : sans remettre en cause le rôle des gènes dans les mécanismes évolutionnaires et en tant qu'unité de sélection naturelle, de nombreux philosophes et biologistes[38] se sont en revanche élevés contre l'idée que seuls les gènes interviendraient dans les mécanismes évolutionnaires et que seuls les gènes pourraient être considérés comme des unités de sélection, une idée notoirement défendue sous le nom de sélectionnisme génétique, par Richard Dawkins (Dawkins, 1976).

Enfin, en ce qui concerne l'explication physiopathologique de la maladie, nous avons montré dans les chapitres 1 et 2, comment on était passé d'une explication génétique des maladies mendéliennes monogéniques, en s'appuyant sur le dogme central de la biologie moléculaire pour proposer l'équation un gène muté – une protéine mutée – une maladie, à des explications plus complexes des maladies monogéniques d'une part et à des explications qui concernent les maladies

[38] Pour une présentation synthétique de ces deux débats, on se reportera à l'analyse de Thomas Pradeu (Pradeu, 2011). Pour une revue de littérature sur le débat des unités de sélection, on se reportera à l'analyse de Sober (Sober et Wilson, 1994).

polygéniques d'autre part. Il est donc bien évident que le gène est une partie de l'*explanans* des explications physiopathologiques des maladies mais qu'un *explanans* complet ne saurait faire l'économie des déterminants épigénétiques et environnementaux des maladies.

Si nous soulignons que les gènes sont une partie de l'*explanans* des explications épidémiologique, évolutionnaire et physiopathologique de la maladie, ce n'est pas dans le but de revenir à une forme d'exceptionnalisme génétique qui ferait du gène un niveau privilégié ou causalement prédominant dans l'explication des maladies. Il serait tout aussi valide de s'intéresser à l'existence éventuelle d'une théorie épigénétique des maladies, puisque l'épigénétique tend à devenir également une partie de l'*explanans* dans les explications épidémiologique, évolutionnaire et physiopathologique des maladies. Nous voulons simplement souligner que cette caractéristique augure de l'intérêt d'une théorie médicale partielle, centrée sur le rôle des gènes dans l'explication des maladies.

4.5.2. Le gène, concept en tension, qui permet le pluralisme explicatif

À ce stade de notre argumentation, on pourrait cependant nous objecter qu'il peut paraître étonnant de s'intéresser au gène comme fondement d'une théorie médicale partielle, alors même que plusieurs philosophes et biologistes ont prédit la disparition du concept de gène, au motif que celui-ci n'a pas d'unité conceptuelle réelle. Si nous avons montré que les gènes interviennent dans les explications épidémiologique, évolutionnaire et physiopathologique des maladies, ce n'est d'ailleurs probablement pas le même concept de gène qui est utilisé dans ces trois types d'explications.

Quelle est la nature exacte du débat ? Nous ne reprendrons pas en détail l'histoire du concept de gène qui a été abondamment étudiée[39], mais on considère généralement que trois grandes conceptions se sont succédées depuis l'invention du terme en 1909 : une conception classique du gène essentiellement développée dans les années 1910-1930 où le gène est avant tout une unité de transmission de l'hérédité mendélienne, une conception moléculaire du gène qui a émergé autour du

[39] Pour une histoire détaillée du concept de gène, voir les travaux de Portin (Portin, 2002) et l'ouvrage de Jean Deutsch (Deutsch, 2012).

développement de la biologie moléculaire et qui identifie la structure d'un gène en tant que segment d'ADN codant à sa fonction, la production d'une protéine faite d'acides aminés par le biais de l'ARN messager. Cependant, à partir des années 1970, de nombreuses découvertes comme celles du mécanisme d'épissage alternatif, de la pléiotropie, des séquences promotrices, des gènes *enhancers* et *silencers*, de l'ADN non codant, pour n'en citer que quelques-unes, bouleversent la correspondance entre structure et fonction.

Nombreux sont les auteurs qui ont mis l'accent sur cette dissolution du concept de gène (Burian, 1985, 2004, 2013 ; Falk, 2000 ; Gayon, 2007 ; Griffiths et Neumann-Held, 1999), ainsi que sur la persistance dans les différentes disciplines biologiques de plusieurs concepts de gène distincts. Ainsi, Griffiths et Stotz distinguent trois concepts de gène (Griffiths et Stotz, 2006) : le « gène instrumental », le « gène nominal » et le « gène post-génomique ». Le gène instrumental est avant tout un instrument formel qui constitue une unité de transmission des caractères mendéliens et qui est donc fortement utilisé par les biologistes évolutionnistes et par la génétique des populations. Le gène nominal est véritablement un héritage de la biologie moléculaire : il est défini par la linéarité de sa séquence, c'est-à-dire par les outils du génie génétique qui ont permis de l'identifier et de l'annoter. Le gène post-génomique est l'héritier des découvertes des années 1970 et du Projet Génome Humain. C'est aussi le gène qui n'est plus spécifiquement déterminé par sa structure nucléotidique, mais dont la définition est étroitement liée au phénotype protéique qu'il peut produire dans un contexte précis et déterminé, « quelque chose que l'organisme peut faire avec son génome ». Dans une étude empirique, Griffiths et Stotz ont montré que ces différents concepts de gène étaient utilisés de façon plus ou moins préférentielle selon les sous-disciplines biologiques (Stotz *et al.,* 2004).

Face à cette crise du concept de gène, plusieurs solutions ont été proposées : certains philosophes proposent de légiférer sur un usage spécifique et restreint du concept de gène, d'autres suggèrent d'abandonner tout à fait le concept de gène. Une troisième voix enfin consiste à assumer le concept de gène dans ses ambiguïtés, soit pour des raisons pragmatiques (c'est l'ambiguïté même du concept de gène qui permettrait aux sous-disciplines biologiques de communiquer entre elles), soit pour des raisons qui tiennent à la thèse du pluralisme explicatif :

> « Ainsi, je propose un pluralisme athéorique similaire à celui que je défends pour les espèces : un gène est n'importe quel bout d'ADN que quiconque a une raison de nommer et de suivre. Les gènes peuvent être des parts propres d'autres gènes ; ils peuvent se chevaucher ; ils peuvent avoir des parties non contiguës, peut-être sur deux ou plusieurs chromosomes. Et, comme c'est illustré dans le cas des gènes responsables des défauts du développement, « gène » peut même référer à une classe de segments d'ADN fonctionnellement connectés. Ma conclusion est qu'il y a des gènes – un point important étant donné à quel point les gens parlent d'eux – mais que le prix à concéder est d'accepter qu'il ne faut pas grand chose pour être un gène. Pas grand chose, mais pas rien non plus. Je fais l'hypothèse que les gènes sont des entités matérielles. » (Dupré, 2012, p. 112)

Ce point est important : dans notre acception du terme de « théorie génétique », le fait que le concept de gène puisse avoir plusieurs définitions selon le type d'explications dans lesquelles il est employé n'est pas un problème, mais une conséquence logique du pluralisme explicatif que nous défendons.

4.5.3. Critères et typologies des théories génétiques possibles

À présent que nous avons justifié doublement l'hypothèse d'une théorie génétique, comme réponse à la généticisation des maladies et comme théorie partielle permettant de combiner différentes explications de la maladie du point de vue du gène, nous pouvons appliquer les critères que nous avons identifiés pour définir une théorie génétique *a minima*.

Nous avons d'abord souligné tout au long du chapitre 3 et dans cette dernière sous-partie une condition importante de toute théorie génétique pour éviter l'écueil du génocentrisme : le cadre d'une théorie génétique doit être celui d'un interactionnisme co-constructionniste. Nous avons défini cet interactionnsime co-constructionniste comme une forme d'interactionnisme qualitatif qui s'appuie sur la thèse de la parité causale et ne donne pas de prédominance causale aux gènes, mais qui distingue au moins d'un point de vue heuristique les facteurs génétiques des facteurs non-génétiques.

La deuxième condition découle directement de notre définition d'une théorie médicale *a minima* : il peut y avoir des *explications génétiques* de la maladie individuelle en tant que maladie-instance, mais il ne peut pas y avoir de *théorie génétique* de la maladie-instance. Une explication génétique de la maladie individuelle serait par exemple : Mme X, âgée de 57 ans, fumeuse depuis vingt ans, qui prend un traitement hormonal substitutif (THS) depuis cinq ans, qui vit dans des conditions socio-économiques précaires, est atteinte d'un cancer du sein, dont la mutation spécifique qu'elle porte sur le gène BRCA1 est une cause parmi d'autres (l'âge, le tabac, le THS, le niveau de vie). Cette mutation spécifique du gène BRCA1 implique qu'on utilise telle stratégie thérapeutique étant donné qu'on sait que les cancers du sein qui présentent cette mutation spécifique, ont tendance à évoluer de telle manière et à être plus sensible à certaines stratégies thérapeutiques. Mais, en aucun cas, cette explication génétique du cancer du sein de Mme X, qui est une explication clinique qui instancie une partie du réseau causal du cancer du sein et qui intègre plusieurs éléments des explications évolutionnaire, épidémiologique et physiopathologique du cancer du sein, ne peut prétendre au statut de « théorie génétique », puisqu'elle explique une maladie individuelle en tant que maladie-instance.

En revanche, à partir du moment où une théorie explique (soit par une explication épidémiologique, soit par une explication évolutionnaire, soit par une explication physiopathologique) le rôle des facteurs génétiques dans une des cibles explicatives (causes, symptômes, évolution, traitement) d'une maladie en tant que maladie-type ou d'une classe de maladie ou de la maladie en général, alors on a une théorie génétique *a minima*.

À partir de cette définition *a minima* d'une théorie génétique, il est possible d'imaginer un large spectre de théories génétiques faisant appel à différents types d'explications des maladies et adoptant un niveau de généralité variable (théorie d'une maladie-type, d'une classe de maladie ou de la maladie en général) et dont les conséquences sur notre conception de la maladie seront également différentes. Enfin, pour évaluer la précision d'une théorie génétique donnée, quel que soit son niveau, on peut recourir aux trois critères supplémentaires que nous avons proposés : la théorie génétique explique-t-elle la variabilité interindividuelle, la variabilité intra-individuelle, la comorbidité ? Nous avons résumé l'ensemble de ces critères dans la Figure 19.

Condition de possibilité		Théorie partielle centrée sur le rôle des gènes au sein d'une explication interactionniste co-constructionniste des maladies
Critères d'identification	**Objet**	La maladie
		Une classe de maladie
		Une maladie-type
	Niveau	Théorie génétique générale : *théorie génétique de la maladie* ou *théorie génétique des maladies*
		Théorie génétique régionale
		Théorie génétique d'une maladie
	Cibles explicatives	Causes
		Symptômes
		Évolution
		Traitement
	Type d'explication	Explication épidémiologique
		Explication évolutionnaire
		Explication physiopathologique
Critères d'évaluation		Variabilité interindividuelle
		Variabilité intra-individuelle
		Comorbidité

Figure 19 : Condition de possibilité, critères d'identification et critères d'évaluation d'une théorie génétique. On peut parler d'une théorie génétique *a minima* à partir du moment où une théorie centrée sur le rôle des gènes dans une explication interactionniste co-constructionniste des maladies. Une théorie génétique mobilise au moins un type d'explication (épidémiologique, évolutionnaire, physiopathologique) pour rendre compte au moins d'une classe explicative (causes, symptômes, évolution, traitement) d'un objet-maladie (la maladie, l'ensemble des maladies, une classe de maladie, une maladie-type). Une théorie qui rend compte de l'explication de toutes les maladies est une théorie génétique générale et peut être soit une *théorie génétique des maladies*, un ensemble hétérogène de théories génétiques régionales, chaque théorie génétique étant spécifique d'une classe de maladie, soit une *théorie génétique de la maladie*, c'est-à-dire une théorie générale qui repose sur une définition générale de la maladie qui unifie le rôle causal commun des gènes. Trois critères (explication de la variabilité interindividuelle, de la variabilité intra-individuelle, de la comorbidité) permettent d'évaluer le degré de précision d'une théorie génétique donnée à un niveau donné.

Il peut donc exister un grand nombre de théories génétiques différentes, on peut ainsi imaginer qu'il existe, voire même qu'il coexiste des théories génétiques de maladies individuelles, des théories génétiques spécifiques à chaque classe de maladie et une théorie génétique de la maladie en général. Étant donné que notre objectif n'est pas de comprendre en quoi une maladie-type peut être dite « génétique », mais de comprendre en quel sens toutes les maladies sont génétiques (la généticisation non pas d'une maladie mais de toutes les maladies), nous allons nous concentrer sur l'analyse de deux formes particulières de théories génétiques *générales* parmi le spectre des théories génétiques possibles : une théorie génétique *des maladies* et une théorie génétique *de la maladie.*

Pour bien comprendre pourquoi nous isolons ces deux formes particulières de théories génétiques et leurs conséquences, repartons de la description de notre théorie médicale contemporaine selon Thagard. Thagard considère que notre théorie médicale contemporaine est un ensemble hétéroclite de théories régionales, c'est-à-dire de théories de classes de maladie. Parmi ces théories régionales, Thagard considère qu'il y a une théorie génétique qui s'applique à une classe particulière de maladies (les maladies génétiques) et qui est donc une théorie des maladies génétiques. La première partie de cette thèse s'inscrit en faux contre cette représentation des maladies génétiques comme une classe à part entière de maladies. Ce que suppose la généticisation, c'est que les gènes ont un rôle causal commun à toutes les maladies. Nous nous proposons donc d'explorer un changement de hiérarchie dans le schéma explicatif de Thagard, en passant d'une théorie génétique régionale supposée (la théorie des maladies génétiques) à l'hypothèse d'une théorie génétique générale (Figure 20).

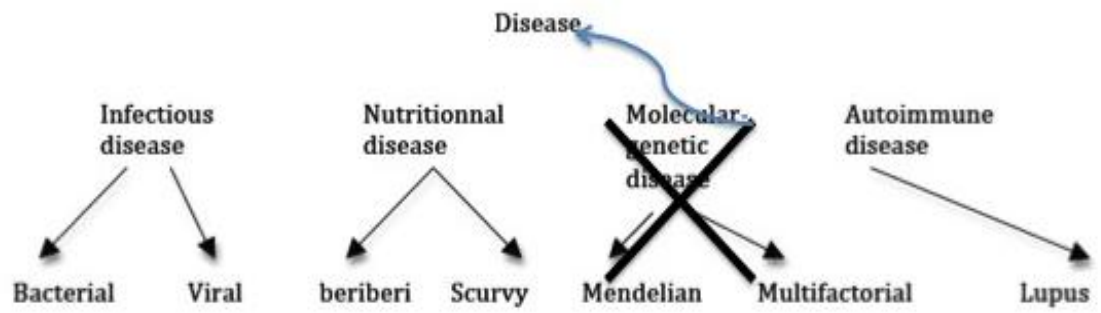


Figure 20 : Passer d'une théorie des maladies génétiques à une théorie génétique générale - (Figure modifiée à partir de Thagard, 2000, p. 35). Plutôt que de considérer les maladies génétiques comme une classe de maladie parmi d'autres (niveau intermédiaire), nous faisons l'hypothèse qu'il y a une théorie génétique générale (niveau supérieur).

Nous avons préalablement identifié deux formes de théorie générale : une théorie générale *des maladies* et une théorie générale *de la maladie*. La première forme s'apparente à notre théorie médicale contemporaine, c'est une théorie générale des maladies, c'est-à-dire une juxtaposition de théories régionales distinctes, chaque théorie s'appliquant spécifiquement à une classe de maladie. Appliqué au concept d'une théorie partielle, cela signifierait que pour chaque classe de maladie telles que nous les connaissons, nous pouvons identifier des théories génétiques régionales qui unifient chacune de ces classes du point de vue génétique (voir Figure 21).

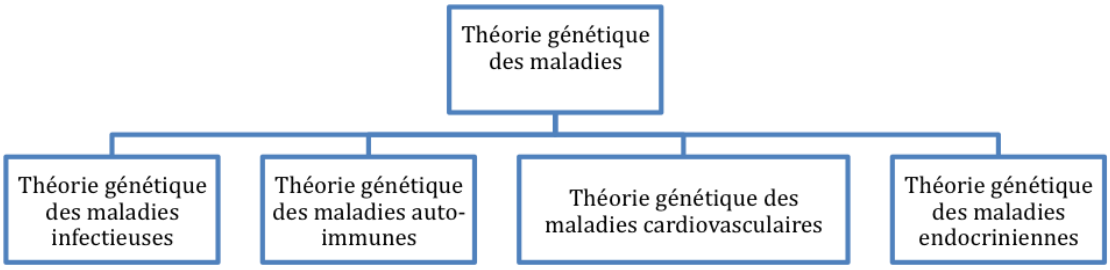

Théorie génétique des maladies
Théorie génétique des maladies infectieuses
Théorie génétique des maladies auto-immunes
Théorie génétique des maladies cardiovasculaires
Théorie génétique des maladies endocriniennes

Figure 21 : Représentation typique d'une théorie génétique des maladies : La théorie génétique des maladies est un ensemble de théories génétiques régionales. Pour chaque classe de maladie, il existe une théorie génétique avec des mécanismes génétiques spécifiques. Ces mécanismes génétiques peuvent varier d'une classe de maladie à une autre. Ce type de théorie ne modifie pas la façon dont on classe les maladies.

Dans cette théorie génétique des maladies, chaque classe de maladie a sa propre théorie génétique régionale, sans que ces théories régionales n'aient de principes ou de mécanismes communs. De façon remarquable, une théorie génétique des maladies ne suppose donc pas nécessairement que l'on modifie notre classification des maladies.

La seconde forme de théorie générale est une théorie générale de la maladie : semblable dans sa forme à la théorie hippocratique, elle repose sur une définition de la maladie et du rôle de plusieurs facteurs causaux dans la maladie (les quatre humeurs), à partir de laquelle toute théorie d'une maladie type peut être dérivée. Appliqué au concept d'une théorie partielle génétique, cela signifierait qu'on dispose d'une explication du rôle commun que jouent les gènes dans toutes les maladies et que c'est à partir de cette explication que peuvent être dérivées l'ensemble des théories génétiques de chaque maladie-type. Il est intéressant de noter que dans une telle théorie, il est tout à fait possible que les classes de maladie telles que nous les connaissons soient amenées à changer, ou pour le moins, qu'une classification des maladies du point de vue de la génétique ne coïncide pas avec notre classification usuelle des maladies (voir Figure 22).

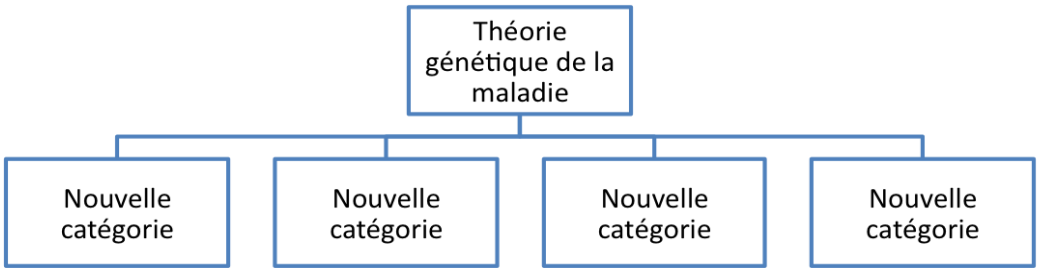


Figure 22 : Représentation typique d'une théorie génétique de la maladie. Une théorie génétique de la maladie reposerait sur une définition du rôle commun des gènes dans la maladie. En fonction de cette définition, des principes de classification peuvent permettre de redéfinir la façon dont nous classons les maladies.

4.6. Conclusion du chapitre : la généticisation révèle-t-elle une théorie génétique des maladies ou une théorie génétique de la maladie ?

L'objectif de ce chapitre était de clarifier ce que nous entendions par le terme de « théorie génétique », en adoptant une position que nous avons qualifiée de « minimalement normative », c'est-à-dire qui consiste à définir les conditions de possibilité d'une théorie génétique *a minima*, tout en envisageant le spectre des formes possibles que pourrait prendre une théorie génétique et des critères qui

puissent permettre d'évaluer le degré de précision d'une théorie génétique donnée, voire de comparer entre elles des théories.

Nous avons défini les théories médicales comme des explications de la maladie pouvant avoir quatre objets différents (la maladie-instance, la maladie-type, la classe de maladie, la maladie) correspondant à différents niveaux de généralité (théorie d'une maladie, théorie régionale de la maladie, théorie générale de la maladie ou théorie générale des maladies) mais visant les mêmes cibles explicatives (causes, symptômes, évolution, traitement) de la maladie, quoique par des modes explicatifs différents (explication épidémiologique, évolutionnaire, physiopathologique). À partir de cette grille analytique et de la définition d'une théorie scientifique comme l'explication de classes de faits indépendants, il nous a été possible de définir une théorie médicale *a minima* et trois classes de faits indépendants (variabilité interindividuelle, variabilité intra-individuelle, comorbidité) dont l'explication augmente la précision d'une théorie.

Nous avons alors pu définir le concept de théorie génétique comme une théorie médicale partielle qui se focalise sur le gène comme *explanans* dans la maladie et définir une théorie génétique *a minima* comme une théorie qui explique (soit par une explication épidémiologique, soit par une explication évolutionnaire, soit par une explication pathophysiologique) le rôle des facteurs génétiques dans une des cibles explicatives (causes, symptômes, évolution, traitement) d'une maladie en tant que maladie-type, ou d'une classe de maladie, ou de la maladie en général. À partir de cette définition d'une théorie génétique *a minima*, et en gardant à l'esprit la nécessité de l'interactionnisme co-constructionniste pour proposer une explication ni génocentriste ni triviale du rôle commun des gènes dans les maladies, il est possible d'envisager un spectre de théories génétiques possibles.

Nous avons souligné en particulier les conséquences de deux formes de théories génétiques : une théorie génétique des maladies serait un ensemble hétérogène de théories génétiques régionales, sans conséquence sur nos classifications des maladies, tandis qu'une théorie génétique de la maladie serait une théorie reposant sur une définition partielle de la maladie en fonction des facteurs génétiques et à partir de laquelle toute théorie d'une maladie-type pourrait être

instanciée, remettant potentiellement en cause la façon dont nous classons les maladies.

Dans la troisième partie de ce travail, nous allons à présent confronter cette conception minimalement normative des théories génétiques à deux exemples de théories génétiques issus de la littérature biomédicale contemporaine : la théorie génétique des maladies infectieuses d'une part (chapitre 5) et la théorie de la médecine des réseaux d'autre part (chapitre 6).

Partie 3 : Deux exemples contemporains de théorie génétique

Dans la première partie de ce travail, à la fois descriptive et critique, nous nous sommes confrontée aux problèmes scientifiques et philosophiques que pose la définition du concept de maladie génétique, pour conclure qu'il fallait abandonner le concept de maladie génétique et le problème de la sélection causale, mais prendre au sérieux la généticisation des maladies. Notre hypothèse est que la généticisation des maladies refléterait l'élaboration d'une théorie génétique de la maladie.

La deuxième partie de ce travail, que nous avons qualifiée de « minimalement normative », nous a permis de raffiner cette hypothèse et d'établir une série de conditions de possibilité d'une théorie génétique de la maladie. Nous avons identifié deux types de conditions : 1) une condition spécifique au caractère génétique de la théorie : l'intégration de cette théorie génétique à une explication interactionniste co-constructionniste des maladies et 2) une série de conditions spécifiques à une théorie médicale qui se focaliserait sur un composant causal de la maladie parmi d'autres (les gènes en l'occurrence), c'est-à-dire une série de conditions spécifiques à une théorie médicale partielle. Notre hypothèse initiale a donc à présent la forme d'une alternative : soit la généticisation des maladies reflète *une théorie génétique des maladies*, c'est-à-dire un ensemble de théories génétiques régionales, spécifiques de chaque classe de maladie telles que nous les connaissons, soit la généticisation des maladies reflète une *théorie génétique de la maladie*, c'est-à-dire l'identification du rôle causal commun des gènes dans toutes les maladies, et cette théorie est susceptible non seulement d'être instanciée pour chaque maladie-type mais aussi de renouveler en profondeur notre classification des maladies.

Cette troisième partie sera donc expérimentale[40] : il s'agit à présent de tester notre hypothèse, en identifiant des tentatives plus ou moins explicites de théories

[40] Précisons que le terme de « partie expérimentale » ne renvoie nullement au mouvement de la philosophie expérimentale qui teste des intuitions philosophiques ou des notions du sens commun par des expériences de sciences cognitives ou de psychologie expérimentale. Il ne s'agit pas ici de tester si notre concept de théorie génétique correspond aux intuitions des individus sur ce concept. Notre démarche n'est pas non plus empirique au sens où nous n'appuyons pas notre démonstration sur une enquête sociologique ou linguistique. Il ne s'agit pas de recenser par des méthodes quantitatives les occurrences du terme « théorie génétique » ou d'analyser les contextes socioculturels dans lesquels ce

génétiques dans la littérature biomédicale contemporaine et en confrontant ces exemples aux conditions nécessaires que nous avons identifiées pour pouvoir parler de théorie génétique *des maladies* ou de théorie génétique *de la maladie*.

Cette troisième partie expérimentale soulève cependant trois difficultés. La première est que la littérature biomédicale contemporaine est littéralement submergée par la génétique. Une simple recherche sur la base de données Pubmed avec les termes « *genetic* » et « *disease* », renvoie ainsi à 268 494 résultats le 13 janvier 2014, le terme « *genetic disease* » à la même date donne 4 777 résultats. Plus précisément, la littérature biomédicale contemporaine est submergée par des articles traitant de l'architecture génétique d'une maladie individuelle, alors que les articles qui cherchent à extrapoler les connaissances acquises sur le rôle causal des gènes à une classe de maladie ou à la maladie en général sont beaucoup moins nombreux et plus difficiles à identifier. Dans un tel contexte, il est impensable de parvenir à une analyse exhaustive de la littérature biomédicale contemporaine : nous nous proposons donc de ne tester qu'un seul exemple de théorie génétique des maladies et un seul exemple de théorie génétique de la maladie.

Cette première difficulté nous amène naturellement à un second problème : comment, dans ce contexte, tester l'hypothèse d'une théorie génétique des maladies, qui semble supposer de rechercher pour chaque classe de maladie s'il existe une théorie génétique correspondante ? Une telle analyse déborderait largement les limites de ce travail. Nous nous proposons donc de choisir un exemple de théorie génétique régionale – la théorie génétique des maladies infectieuses, un des rares exemples de théorie génétique explicite que l'on peut trouver dans la littérature biomédicale contemporaine et d'utiliser cet exemple pour analyser ce qu'est une théorie génétique régionale et pour envisager à quoi ressemblerait une théorie génétique des maladies à partir de cet exemple.

terme émerge. Nous considérons que cette partie est « expérimentale », au sens où nous avons formulé une hypothèse sur le statut épistémique de la généticisation que nous testons à présent en confrontant les critères de deux types de théories génétiques possibles à un corpus d'articles scientifiques. Il s'agit de démontrer qu'il existe des théories génétiques dans la littérature biomédicale contemporaine et de caractériser les phénomènes qu'elles permettent d'expliquer.

La troisième difficulté est qu'une recherche PubMed ne nous a pas permis d'identifier une théorie génétique de la maladie, qui se présente explicitement comme telle : comment tester alors l'hypothèse d'une théorie génétique de la maladie ? Pour trouver une théorie de ce type, nous nous sommes tournée vers des programmes de recherche qui s'éloignaient autant que possible des explications individuelles des maladies pour s'intéresser aux principes généraux d'explications de la maladie, c'est-à-dire vers le cadre théorique de la biologie des systèmes et de la médecine des systèmes. Ces nouvelles traditions de recherche en biologie et en médecine ont l'ambition de combiner des outils de modélisation et l'étude des mécanismes moléculaires du vivant, pour comprendre les principes d'organisation du vivant et les principes généraux de la maladie. C'est dans ce contexte que nous avons identifié un programme de recherche particulier, la médecine des réseaux, comme un cadre possible pour l'émergence d'une théorie génétique de la maladie. En effet, la médecine des réseaux ne s'intéresse pas à l'explication de la génétique de chaque maladie : elle prétend comprendre quels principes généraux sur l'architecture des maladies nous pouvons tirer des connaissances que nous avons de l'explication génétique des maladies individuelles.

Cette partie expérimentale ne prétend donc certainement pas recenser exhaustivement l'ensemble des théories génétiques régionales et générales qui sont possiblement en train d'émerger au sein de la littérature biomédicale contemporaine. Elle consiste seulement à discuter deux exemples de théories génétiques : d'une part, un exemple explicite de théorie génétique régionale, la théorie génétique des maladies infectieuses, à partir duquel on peut imaginer ce que serait une théorie génétique *des maladies* (chapitre 5) et d'autre part, un exemple implicite – la médecine des réseaux – qui est issu de la tradition de recherche de la biologie et de la médecine des systèmes et qui nous semble être un cadre fécond pour imaginer ce à quoi ressemblerait une théorie génétique *de la maladie* (chapitre 6). Pour chacun de ces exemples, nous montrerons en quoi il respecte ou non les contraintes que nous avons identifiées dans la deuxième partie de cette thèse, et en quoi il renouvelle notre explication et notre compréhension du rôle causal commun des gènes dans toutes les maladies.

Au chapitre 5, nous examinerons en quoi la théorie génétique des maladies infectieuses est un modèle pour penser une théorie génétique des maladies. L'hypothèse d'une explication génétique des maladies infectieuses a émergé comme une réponse au problème de la variabilité interindividuelle clinique aux infections, insuffisamment expliquée par la théorie des germes, la théorie microbienne, la théorie immunologique et la théorie écologique des maladies infectieuses. Très rapidement, deux traditions de recherche, la génétique des populations d'une part et la génétique clinique d'autre part, ont permis de découvrir deux types de mécanismes susceptibles d'expliquer cette variabilité interindividuelle clinique aux maladies infectieuses : l'existence d'une prédisposition polygénique aux infections communes d'une part et l'existence d'une prédisposition mendélienne aux infections multiples (aussi appelée immunodéficience primaire) d'autre part. Mais il a fallu attendre les développements des techniques de génétique moléculaire dans les années 1970 pour découvrir l'existence de deux autres mécanismes, la prédisposition mendélienne à une seule infection d'une part et le concept de gène majeur de prédisposition ou de résistance à une seule infection d'autre part, pour qu'un petit groupe de chercheurs français rattachés au Laboratoire de Génétique des Maladies Infectieuses à l'Hôpital Necker affirme l'existence d'une « théorie génétique unifiée » des maladies infectieuses . Parce qu'elle concerne les maladies infectieuses, exemples paradigmatiques de maladies pensées d'abord comme purement environnementales, cette théorie intervient d'emblée dans une explication interactionniste co-constructionniste des maladies qui identifie une matrice causale complexe. Cette théorie est une théorie partielle des maladies infectieuses, puisqu'elle se focalise sur le rôle des mécanismes génétiques dans l'explication de la variabilité clinique interindividuelle aux maladies infectieuses, mais qu'elle a un rôle complémentaire de la théorie des germes, de la théorie microbienne, de la théorie immunologique et de la théorie écologique dans l'explication des maladies infectieuses. C'est une théorie régionale supposée s'appliquer à l'ensemble des maladies infectieuses et qui contribue à l'explication des quatre cibles explicatives (causes, symptômes, évolution, traitement) des maladies infectieuses en déployant une explication physiopathologique de type mécaniste, mais qui s'appuie et complète des explications épidémiologique et évolutionnaire des maladies infectieuses. Cette théorie génétique des maladies infectieuses a pour mérite

principal d'expliquer la variabilité interindividuelle clinique aux maladies infectieuses, en particulier le taux élevé de mortalité infantile dans les infections primaires, de compléter l'explication de la variabilité interindividuelle des adultes aux infections secondaires et aux réactivations et d'expliquer le problème des infections inapparentes. Ceci dit, la théorie génétique des maladies infectieuses présente aussi quelques limites, dont l'une en particulier peut être un frein à l'extrapolation de ces mécanismes à une théorie régionale génétique des maladies, les définitions de concepts de « prédisposition mendélienne » ou de « prédisposition polygénique » étant fortement remises en question aujourd'hui. On peut cependant imaginer que les autres classes de maladie (maladies auto-immunes, cancers, etc.) admettent des théories génétiques régionales spécifiques à chaque classe de maladie.

Au chapitre 6, nous examinons une approche particulière appelée « médecine des réseaux », essentiellement développée par l'équipe d'Albert-László Barabási et qui nous paraît le cadre le plus intéressant pour une *théorie génétique de la maladie*. À l'intersection de la naissance du concept de « gènes humains de la maladie », du développement de la biologie et de la médecine des systèmes et de la formalisation de la théorie des réseaux, la médecine des réseaux n'étudie pas les maladies comme des entités physiopathologiques distinctes mais comme des phénomènes partageant des réseaux d'éléments biologiques en commun dont l'organisation normale est rompue (Barabási *et al.,* 2011).

Au sein de la médecine des réseaux, une attention particulière est portée aux réseaux de gènes humains de la maladie, dont on présuppose qu'ils ont des propriétés différentes des autres gènes humains. Ces propriétés sont explorées par la construction et l'analyse du « diseasome », un graphique bipartite permettant de visualiser toutes les associations gènes-maladies, à partir duquel on peut construire d'une part le réseau des gènes humains de la maladie et d'autre part le réseau des maladies humaines. Dans le réseau des gènes humains de la maladie, les gènes sont les nœuds du réseau et deux gènes sont reliés s'ils sont impliqués dans la même maladie. Dans le réseau des maladies humaines, les maladies sont les nœuds du réseau et deux maladies sont reliées si elles partagent un gène commun dans leur physiopathologie.

Trois analyses peuvent être tirées du diseasome. (1) La première est une analyse globale qui permet de caractériser la structure générale du réseau. (2) La seconde est une analyse locale, qui consiste à recouper l'analyse des propriétés topologiques des réseaux des gènes humains de la maladie avec des informations sur leur rôle biologique ou physiopathologique. (3) La troisième consiste à comparer les propriétés topologiques des réseaux des gènes humains de la maladie en les comparant aux propriétés des réseaux de l'interactome, qui est supposé désigner l'ensemble des interactions biologiques possibles dans un être humain. En se fondant sur ces trois analyses, la médecine des réseaux peut ainsi développer une définition générale de la maladie qui s'appuie sur le concept de module fonctionnel de la maladie. Selon cette définition, la maladie n'est jamais le résultat du défaut d'un seul gène ou d'un seul composant de l'organisme, elle est le résultat de perturbations des éléments composant son module fonctionnel.

Dans ce chapitre, nous soutenons donc que la médecine des réseaux est une théorie générale de la maladie qui repose sur une définition modulaire de la maladie. Cette définition générale de la maladie porte intrinsèquement la marque d'un interactionnisme co-constructionniste qui répond aux critères que nous avons identifiés dans le chapitre 3. Par ailleurs, au sein de ce cadre interactionniste co-constructionniste des maladies, il est possible de définir un rôle causal commun pour les gènes dans l'explication des maladies et d'isoler une théorie génétique de la maladie. Les propriétés des gènes humains de la maladie permettent ainsi (1) d'expliquer l'origine génétique commune des maladies, qui recouvre l'explication du phénomène de comorbidité, des familles de syndromes et des classes de maladie, (2) de renouveler la distinction entre maladies monogéniques et maladies polygéniques et (3) de prédire l'existence de nouveaux gènes de la maladie.

Pour terminer ce chapitre, nous nous proposons de revenir sur l'exemple des maladies infectieuses – mais au prisme de la médecine des réseaux. Notre objectif est ici triple : il s'agit de (1) comparer l'explication que théorie génétique des maladies infectieuses et médecine des réseaux donnent de la classe des maladies infectieuses, (2) comparer comment ces deux théories intègrent plus ou moins l'existence d'autres explications partielles (le microbiome, la génétique du parasite) des maladies infectieuses, (3) comparer comment ces deux théories traitent les cas-limites de

maladies infectieuses (cancers viro-induits, maladies auto-immunes, etc.). Cette triple comparaison permet de montrer sur un exemple précis comment la médecine des réseaux permet d'expliquer les maladies infectieuses de façon plus globale (c'est à dire en prenant en compte l'ensemble des facteurs causaux impliqués dans les maladies infectieuses) et plus intégrée (c'est-à-dire en permettant de concilier beaucoup plus facilement les différentes explications des maladies infectieuses) que ne le faisait la théorie génétique des maladies infectieuses. Mais cette triple comparaison permet aussi de comprendre en quel sens la théorie génétique des maladies infectieuses et l'approche des maladies infectieuses par la médecine des réseaux ne sont pas des explications contradictoires, mais bien des explications complémentaires des maladies infectieuses. Autrement dit, la pluralité des théories génétiques n'est pas un problème, mais une nécessité.

Chapitre 5 : La théorie génétique des maladies infectieuses, un exemple de théorie génétique régionale

Dans la présentation de cette troisième partie, nous avons justifié notre choix d'examiner la théorie génétique des maladies infectieuses parce que c'est l'un des rares exemples de théorie génétique explicite d'une classe de maladie que l'on trouve dans la littérature biomédicale contemporaine. Mais en dehors de cette justification méthodologique, plusieurs arguments font de la théorie génétique des maladies infectieuses un exemple particulièrement intéressant pour notre propos.

Premièrement, les maladies infectieuses sont souvent pensées comme l'exemple paradigmatique de maladies purement environnementales, dont la cause principale serait un agent infectieux, c'est-à-dire un micro-organisme pathogène. En examinant la théorie génétique des maladies infectieuses, nous avons donc une preuve supplémentaire de l'extension de la généticisation à toutes les maladies et de la nécessité de repenser le rôle commun des gènes dans toutes les maladies, y compris les maladies infectieuses. Dans le même temps, une théorie génétique des maladies infectieuses est à l'abri, par définition, d'une accusation de génocentrisme et d'un retour au problème de la sélection causale : il ne peut s'agir bien sûr d'affirmer que les gènes sont la cause principale des maladies infectieuses, puisque développer une maladie infectieuse suppose toujours d'avoir été infecté par un agent infectieux (condition nécessaire mais non suffisante).

Deuxièmement, les maladies infectieuses sont aujourd'hui à un carrefour théorique : autrefois dominée par la théorie des germes, l'explication des maladies infectieuses s'est progressivement enrichie des apports de différentes disciplines ou domaines de recherche. Ainsi, les développements de l'immunologie ont permis de comprendre le rôle du système immunitaire dans le développement et l'évolution d'une maladie infectieuse. Les travaux sur la génétique des micro-organismes pathogènes ont permis d'introduire la notion de « virulence » d'un agent infectieux et de comprendre l'influence de la souche infectieuse sur le développement de la maladie. L'écologie explique l'influence des facteurs environnementaux et sociaux dans l'émergence des maladies infectieuses. Plus récemment, plusieurs programmes de recherche comme le *Human Microbiome Project*, une initiative du département

national de santé américain démarrée en 2008 (Turnbaugh *et al.,* 2007), ont mis l'accent sur le rôle du microbiome (l'ensemble de la flore commensale d'un organisme) dans la modulation de la réponse du système immunitaire. Il est donc intéressant de questionner le positionnement de la théorie génétique par rapport à ces autres explications des maladies infectieuses et dans l'explication des maladies infectieuses en général. Autrement dit, l'exemple des maladies infectieuses offre un observatoire privilégié pour comprendre ce qu'une théorie génétique régionale peut apporter à l'explication d'une classe de maladie par rapport aux autres formes d'explications qui coexistent déjà pour cette classe de maladie.

Troisièmement, et ce point est lié au précédent, la théorie génétique des maladies infectieuses a un objectif explicite très précis : celui d'expliquer la variabilité interindividuelle à la maladie infectieuse et plus précisément quatre classes de faits qui sont non expliqués ou incomplètement expliqués par la théorie des germes, « la théorie microbienne », « la théorie immunologique » et la « théorie écologique »[41], à savoir (1) les porteurs asymptomatiques – c'est-à-dire le fait que certains individus infectés ne développent jamais la maladie, (2) la variabilité des formes cliniques d'une même maladie, (3) la variabilité de la sévérité des infections causées par un même pathogène et en particulier l'observation épidémiologique selon laquelle il existe un taux de mortalité élevé des infections primaires infantiles tandis que les infections secondaires et des réactivations chez les adultes ont une mortalité moins sévère, (4) les cas de résistance aux infections.

Nous commencerons par présenter les raisons pour lesquelles la théorie génétique des maladies infectieuses a émergé et le cadre théorique dans lequel cette théorie a émergé. Puis nous présenterons le cœur de la théorie génétique des maladies infectieuses telle qu'elle est exposée par ses théoriciens, Jean-Laurent Casanova et Laurent Abel. En reprenant le cadre conceptuel que nous avons développé dans la deuxième partie de cette thèse, nous expliciterons alors ce que la théorie génétique des maladies infectieuses explique et comment elle l'explique. Plus particulièrement, nous montrerons en quoi la théorie génétique des maladies

[41] Nous mettons entre guillemets les termes de « théorie microbienne », « théorie immunologique » et « théorie écologique » parce que c'est une terminologie non conventionnelle employée par Casanova et Abel. Nous explicitons à quoi ces termes font référence dans la section 5.1. de ce travail.

infectieuses est un exemple de théorie génétique régionale et quelles sont les limites de cette théorie. Enfin, nous examinerons s'il est possible d'extrapoler le contenu de la théorie génétique des maladies infectieuses aux autres classes de maladie, c'est-à-dire que nous chercherons à déterminer si et en quoi la théorie génétique des maladies infectieuses peut servir de modèle à une théorie génétique générale *des maladies*.

5.1. Pourquoi faire l'hypothèse d'une théorie génétique des maladies infectieuses ?

Avant de présenter la théorie génétique des maladies infectieuses, il nous faut dire quelques mots des origines historiques et des raisons pour lesquelles une telle théorie a été développée. Le terme de « théorie génétique des maladies infectieuses » a été introduit[42] par les équipes de deux chercheurs, Laurent Abel (qui a une formation d'épidémiologie génétique) et Jean-Laurent Casanova (qui a d'abord une formation d'immunologie), qui se sont réunis en 2001 pour créer le Laboratoire de Génétique des Maladies Infectieuses de l'Université Necker. La théorie génétique des maladies infectieuses (que nous abrègerons sous le nom de TGMI dans la suite de notre travail) désigne un ensemble de quatre mécanismes génétiques communs aux maladies infectieuses, qui permettent de « définir les variations génétiques qui rendent compte de la variabilité interindividuelle au cours d'une infection » (Casanova et Abel, 2007a, p. 915). La variabilité interindividuelle aux infections est elle-même présentée comme « une caractéristique de la plupart des infections dans les populations humaines du monde entier et à travers l'histoire […], qui va des infections asymptomatiques aux infections létales. » (Casanova et Abel, 2007a, p. 915).

[42] Si le terme de « théorie génétique des maladies infectieuses » apparaît pour la première fois à notre connaissance dans un article de 2005 (Alcaïs, Fieschi, *et al.*, 2005), en référence à un autre article qui n'emploie pourtant pas explicitement ce terme (Casanova et Abel, 2005), on trouve déjà les concepts majeurs de cette théorie dans les premiers articles de Jean-Laurent Casanova et Laurent Abel (Abel et Casanova, 2000), dès leurs débuts au Laboratoire de Génétique des Maladies Infectieuses. Depuis le terme de « théorie génétique des maladies infectieuses » a été explicitement utilisé dans quatre articles majeurs de revues de la littérature sur lesquels nous nous appuierons principalement dans la suite de ce travail (Alcaïs *et al.,* 2009 ; Alcaïs, Abel, *et al.,* 2010 *;* Alcaïs, Quintana-Murci, *et al.,* 2010 ; Casanova et Abel, 2007a, 2013).

5.1.1. La variabilité interindividuelle aux infections, un manque explicatif de la théorie des germes

La catégorie des maladies infectieuses est née avec la théorie des germes, au sens où la théorie des germes permet d'identifier chaque maladie infectieuse à un seul agent pathogène et définit la classe des maladies infectieuses comme la classe des maladies qui sont causées par un micro-organisme pathogène spécifique. La théorie des germes a été longtemps l'objet d'une historiographie souvent dithyrambique et parfois peu critique, tant elle a été associée à l'avènement scientifique de la médecine, à des succès thérapeutiques remarquables avec le développement des mesures d'hygiène, des vaccins et des antibiotiques, et enfin au développement du modèle monocausal de la maladie, avec la reconnaissance de la spécificité de chaque germe à une infection particulière (Broadbent, 2009, 2013 ; Carter, 2003). Si plusieurs auteurs ont insisté sur le fait que le concept de germes et l'hypothèse selon laquelle les germes étaient impliqués dans les maladies étaient déjà discutés dès les années 1840 (Carter, 2003 ; Worboys, 2000, 2007),on considère que la théorie des germes, aussi appelée « théorie microbienne » fait principalement référence aux travaux de Louis Pasteur et de Robert Koch entre les années 1860 et les années 1880, avec la réfutation de l'hypothèse de la génération spontanée, la mise en évidence du rôle des germes dans la fermentation, de l'existence ubiquitaire des germes, la découverte du bacille de Koch dans la tuberculose et la mise au point des premiers vaccins. La théorie des germes est souvent associée aux postulats d'Henlé-Koch[43], qui sont moins un ensemble de critères rigides et nécessaires pour affirmer la cause d'une maladie infectieuse qu'un ensemble de critères expérimentaux qui guident le travail

[43] Le nom de ces postulats et leur appartenance a été très discuté. « Je dirais que la Révolution Bactériologique a été rapide – concentrée dans les années 1880 et qu'elle fut radicale – elle a « transformé tous les aspects de la médecine ». On pourrait considérer que l'annonce par Robert Koch de la découverte du bacille de la tuberculose le 22 mars 1882 aura été l'équivalent de la prise du Palais d'Hiver en 1917. On a dit que cette découverte avait introduit une compréhension fondamentalement nouvelle des maladies infectieuses et avait démontré avec succès la valeur de ces nouvelles méthodes pour enquêter sur les causes et les processus de la maladie dans le laboratoire. Ce qui a été déterminant, ce n'est pas seulement la découverte d'un micro-organisme comme cause essentielle de la tuberculose, bien que cette découverte ait eu des conséquences drastiques puisque la maladie considérée comme héréditaire est devenue transmissible, mais les nouvelles significations de la maladie et les nouvelles façons de connaître ses causes, ses processus et ses conséquences. Ce changement a été canonisé sous le nom de « Postulats de Koch », bien que son collègue Loeffler les ait formulés le premier et qu'il est discutable que Koch les ait jamais défendus lui-même. » (Worboys, 2007, p. 23)

d'identification du micro-organisme responsable de l'infection par le bactériologiste : (1) le micro-organisme doit être présent en abondance dans tous les organismes souffrant de la maladie, mais absent des organismes sains, (2) le micro-organisme doit pouvoir être isolé de l'organisme malade et cultivé dans des cultures pures, (3) le micro-organisme cultivé doit entrainer l'apparition de la maladie lorsqu'il est introduit dans un organisme sain, (4) le micro-organisme doit être à nouveau isolé du nouvel organisme hôte rendu malade puis identifié comme étant identique à l'agent infectieux original.

Précisons que la validité de ces postulats a été âprement discutée dès leur formulation, puisqu'ils n'ont pu être appliqués qu'à un petit nombre de maladies, comme le reconnaissait Koch lui-même (Evans, 1976, 1991 ; Stewart, 1968). D'une part, le postulat (1) a été remis en cause par la découverte des porteurs asymptomatiques. Les porteurs asymptomatiques désignent les individus qui sont infectés par un agent infectieux mais ne développent pas pour autant la maladie. L'exemple le plus célèbre de porteur asymptomatique est le cas de Mary Mallon, une cuisinière américaine, surnommée Mary Thypoïde, dont on a estimé qu'elle avait involontairement contaminé 51 personnes aux États-Unis, sans jamais avoir développé elle-même les symptômes de la fièvre typhoïde et alors même que son autopsie a permis de retrouver des bactéries de la fièvre typhoïde vivantes dans sa vésicule biliaire (Leavitt, 1996). D'autre part, le postulat (3) a été remis en cause par la découverte de cas de résistance à certaines infections : ainsi certaines prostituées ougandaises, régulièrement exposées au virus du VIH, ne deviennent jamais séropositives. Au contraire des porteurs asymptomatiques, qui sont exposés et infectés mais non malades, elles sont exposées mais non infectées. Un autre exemple célèbre est le cas de Max von Pettenkoffer, un violent opposant à la théorie des germes, qui n'a développé aucun symptôme du choléra, après avoir bu une solution de *Vibrio cholerae*, fournie par Koch lui-même.

De façon plus générale, la théorie des germes ne permet pas de comprendre pourquoi des individus infectés par le même pathogène issu de la même souche (ce n'est donc pas une question de virulence du pathogène) peuvent développer des symptômes ou des formes cliniques différentes. Ainsi, le bacille de la lèpre peut soit provoquer une forme dite paucibacillaire (peu de lésions hypopigmentées et non

sensibles, absence de bactéries à la surface des lésions), soit une forme dite multibacillaire (un grand nombre de lésions douloureuses avec une importante charge bactérienne à la surface). Enfin, comment comprendre que l'évolution d'une même infection puisse être radicalement différente entre deux individus infectés par le même pathogène, c'est-à-dire comment expliquer que des pathogènes peu virulents puissent être responsables d'infections mortelles chez certains individus ? Ce sont ces quatre classes de faits que Casanova et Abel regroupent sous le nom de « variabilité inter-individuelle aux infections » et qu'ils estiment non expliqués par la théorie des germes.

5.1.2. La place des théories microbienne, immunologique et écologique dans l'explication de la variabilité interindividuelle aux infections

5.1.2.1. La « théorie microbienne » n'explique pas la variabilité interindividuelle aux infections lorsque deux individus sont exposés à une même souche pathogène

A ce point il nous faut noter une ambigüité récurrente dans le discours d'Abel et Casanova qui évoquent de façon presque synonyme la théorie des germes telle qu'elle a été proposée à la fin du dix-neuvième siècle et la « théorie microbienne » ou « théorie microbiologique », qui engloberait les évolutions de la théorie des germes au cours du vingtième siècle. En effet, les développements de la microbiologie ont élargi le spectre des maladies infectieuses en identifiant de nouvelles classes de micro-organismes pathogènes (les virus, les parasites, les champignons, les prions) et en identifiant plusieurs facteurs dépendants du micro-organisme pathogène, qui peuvent expliquer une partie de la variabilité interindividuelle aux infections, que ces facteurs soient quantitatifs (avec la notion de « charge » du pathogène, qui désigne le nombre d'agents infectants, la durée de l'exposition au pathogène) ou qualitatifs (comme la notion de virulence). Le point important est que la « théorie microbienne » telle que la définissent Abel et Casanova est une explication des maladies infectieuses centrée sur le rôle du micro-organisme comme cause principale de l'infection :

> « … et ce n'est pas sans raison que la plupart des microbiologistes restent fermement focalisés sur les microbes. De façon bien compréhensible, ils voient

> les microbes comme la cause principale (si ce n'est la seule) des maladies infectieuses, l'acteur principal dans un casting de milliers de figurants, chacun de ces derniers jouant un rôle individuel modeste dans la pathogénèse et agissant comme une niche pour le microbe. » (Casanova et Abel, 2013, 217)

Cette théorie microbienne s'est clairement illustrée par la façon dont « l'école moléculaire » (Méthot, 2012a) a défini les concepts de pathogénicité et de virulence tout au long du vingtième siècle :

> « Les premières définitions de la pathogénicité et de la virulence étaient d'abord centrées sur le pathogène et étaient fondées sur l'hypothèse que ces caractéristiques étaient des propriétés intrinsèques des microorganismes, bien qu'il soit reconnu que la pathogénicité n'était ni invariante ni absolue » (Casadevall et Pirofski, 1999, 3703).

Ainsi, au début du vingtième siècle, le concept de virulence est d'abord défini comme la capacité d'un germe à causer une maladie (paramètre qualitatif), tandis que la pathogénicité désignerait plutôt soit un paramètre de la virulence, soit le pouvoir d'un parasite à envahir les défenses de l'hôte (paramètre quantitatif). Le concept de virulence est donc un critère permettant d'identifier un micro-organisme comme pathogène et de démarquer les micro-organismes pathogènes des non-pathogènes (c'est-à-dire des micro-organismes commensaux par exemple) (Méthot, 2012c). Avec la découverte de la recombinaison génétique chez les bactéries dans les années 1950 par Joshua Ledenberg, puis la découverte que les bactéries pouvaient acquérir des « gènes de virulence » situés dans des « ilots de pathogénicité », le concept de virulence a été de plus en plus identifié à une propriété moléculaire du pathogène, ce qu'illustre bien la notion de postulats moléculaires de Koch proposés par Falkow dans les années 1980 (Falkow, 1988). Les postulats moléculaires de Koch sont un ensemble de critères qui doivent être satisfaits pour démontrer qu'un gène présent dans un micro-organisme pathogène encode un produit qui contribue à la maladie causée par la pathogène, autrement dit ces postulats moléculaires permettent d'identifier « les gènes de virulence ».[44] En résumé, ce que Casanova et Abel appellent « théorie

[44] Les cinq postulats moléculaires définis par Falkow sont : (i) le gène doit être présent dans les souches associées à la pathologie et pas dans les autres ; (ii) la mutation de ce gène doit réduire fortement, sinon

microbienne » désigne moins une théorie unifiée que les différents prolongements de la théorie des germes, c'est-à-dire le développement, en partie à travers les outils de la biologie moléculaire, d'un ensemble de concepts (comme ceux de virulence et de pathogénicité) désignant des propriétés intrinsèques au micro-organisme pathogène, considéré comme la principale cause de l'infection. Cette théorie microbienne explique en partie la variabilité interindividuelle au sens où elle permet de comprendre que des souches différentes d'une même espèce pathogène n'ont pas forcément les mêmes propriétés de virulence et de pathogénicité, ce qui influe sur la sévérité et l'évolution des symptômes de la maladie et ce qui explique qu'il y ait une grande variabilité interindividuelle clinique à l'échelle d'une population infectée par différentes souches. Ceci dit, cette théorie microbienne ne peut expliquer la variabilité clinique interindividuelle (porteurs asymptomatiques, résistance aux infections, variabilité des symptômes, variabilité de la sévérité de l'évolution), lorsqu'une population est confrontée à une même souche d'une espèce pathogène.

5.1.2.2. La « théorie immunologique » n'explique pas les échecs de la vaccination ni les primo-infections infantiles potentiellement mortelles

En parallèle de la théorie des germes et de la théorie microbienne se sont développées d'autres explications complémentaires des maladies infectieuses, notamment ce que Casanova et Abel appellent « la théorie immunologique » et « la théorie écologique ». Si le terme de « théorie des germes » a été consacré par l'usage, la notion de « théorie immunologique » est moins évidente, non pas qu'il n'y ait pas de théories en immunologie, mais plutôt en raison de la multiplication de théories en immunologie, qui ne portent pas toujours sur le même domaine d'extension et qui ont été ou sont parfois encore en concurrence. Ainsi, Anne Marie Moulin montre bien comment la période pastorienne est une période d'empirisme expérimental, une transition permettant de passer des expériences ponctuelles d'immunisation connues déjà dans de nombreux pays et dès l'Antiquité à la science de l'immunologie (Moulin,

abolir, le pouvoir pathogène de ces souches et (iibis) la complémentation en trans doit permettre la restauration du pouvoir pathogène ; (iii) l'expression *in vivo* du gène doit être démontrée ; et (iv) des anticorps spécifiques doivent être produits et conférer une protection, au moins partielle, contre la maladie.

1991). Le début du vingtième siècle voit ensuite s'affronter deux théories sur les défenses du système immunitaire entre les théories de Metchnikoff et d'Ebhert sur l'immunité cellulaire et l'immunité humorale, avec la victoire écrasante de la théorie cellulaire. Autour de la théorie clonale de Mac Burnet s'opère une forme de synthèse dans les années 1960, connue sous le nom de « théorie du soi et du non-soi », qui permet d'établir un critère d'immunogénicité (c'est-à-dire de répondre à la question : qu'est-ce qui déclenche une réponse immunitaire effectrice ?), fondée sur la discrimination du système immunitaire entre le soi et le non-soi. Mais cette théorie, qui s'appuie sur une définition très ambiguë des termes de soi et de non-soi, (Pradeu, 2007) se heurte elle-même à plusieurs découvertes en immunologie (théorie du danger, problème de la tolérance immunologique) qui ont amené Pradeu, Jaeger et Vier à proposer tout récemment de reformuler la théorie immunologique comme une « théorie de la discontinuité » (Pradeu *et al.*, 2013).

Ainsi, il n'y a pas une théorie immunologique, mais des théories immunologiques. Par ailleurs, l'immunologie en général dépasse largement la question des maladies infectieuses : la réponse du système immunitaire aux infections n'est qu'une classe de faits parmi d'autres (la réponse auto-immune, la tolérance fœto-maternelle, etc.) à expliquer en immunologie. Dans les écrits de Casanova et Abel, il semble donc que ce ne soit ni l'unité de l'immunologie, ni une théorie immunologique en particulier qui soit visée : il s'agit plutôt de montrer en quoi l'ensemble des connaissances en immunologie n'expliquent que partiellement la variabilité interindividuelle aux infections. Ainsi, l'immunologie permet de comprendre à la fois la réponse initiale du système immunitaire à une infection (déclenchement immédiat et non spécifique de la réponse immunitaire innée à un antigène considéré comme étranger ou présentant une variation moléculaire non reconnue, puis recrutement d'une immunité adaptative cellulaire ou humorale chargée de lutter plus spécifiquement contre l'agent pathogène) et l'efficacité de l'infection primaire ou de la vaccination pour protéger l'hôte d'une infection secondaire, par la création d'une mémoire immunitaire à long terme qui se déclenchera en première ligne si l'hôte est à nouveau exposé au même agent pathogène ou à un agent pathogène proche sur le plan moléculaire (notion de protection croisée offerte par le vaccin).

En revanche, l'immunologie ne permet pas de comprendre l'échec de la

vaccination chez certains individus, ni le développement chez certains individus de maladies après avoir été exposé à des souches de pathogènes atténués (cas de BCGéites – ces infections potentiellement mortelles développées par certains enfants lorsqu'ils sont vaccinés contre la tuberculose en utilisant le Bacille de Calmette et Guérin, version atténuée du bacille de Koch), ni enfin l'existence d'infections infantiles mortelles chez des enfants considérés comme immunocompétents :

> « Étant donné la solidité de la théorie immunologique des maladies infectieuses, dont les mécanismes ont été élégamment disséqués par les universitaires au cours du dernier siècle et dont les implications pratiques ont été exploitées avec succès par l'industrie, ce n'est pas sans raison que certains immunologistes ne perçoivent pas la nécessité d'une théorie génétique germinale humaine complémentaire. Cependant, l'immunité adaptative est contrôlée génétiquement et un certain nombre d'observations comme l'échec de la vaccination chez certains individus, encouragent une théorie génétique de la réactivation et des infections secondaires comme un complément à la théorie immunologique. Peut-être plus important, la théorie immunologique convient le mieux à la réactivation et aux maladies secondaires et sa portée pourrait bien être limitée à ces maladies. Elle ne peut pas aisément rendre compte de la variabilité interindividuelle au cours des infections primaires naturelles qui arrivent typiquement pendant l'enfance (bien que ce ne soit pas toujours le cas) et contre laquelle la vaccination offre la seule protection (ou, plus rarement, une protection croisée). » (Casanova et Abel, 2013, 218)

5.1.2.3. La « théorie écologique » ne permet pas de comprendre la variabilité interindividuelle aux infections lorsque deux individus sont exposés au même pathogène dans un même environnement

Enfin, la notion de « théorie écologique des infections » ne va pas non plus sans ambiguïté. Tout d'abord, Casanova et Abel sont relativement sibyllins quant à la définition de cette théorie écologique des maladies infectieuses. Ils n'y font allusion que dans leur article consacré à l'histoire de la théorie génétique des maladies infectieuses, en la définissant alors comme la théorie qui explique la variabilité clinique interindividuelle par « la variabilité environnementale (qui n'inclut pas celle du

pathogène considéré mais inclut potentiellement celle d'autres microbes) » (Casanova et Abel, 2013, p. 216). Par ailleurs, si dans la littérature biomédicale scientifique et historique, on trouve parfois les termes de « style écologique » (Méthot, 2012a) ou la notion de « vision écologique de la maladie » (Anderson, 2004), ou encore en anglais les termes d' « écologie des maladies infectieuses » (*infectious diseases ecology*), il n'y a jamais de référence à une « théorie » écologique des maladies infectieuses dans la littérature biomédicale contemporaine.

En revanche, certains chercheurs sont considérés comme des figures de proue d'une vision écologique des maladies infectieuses : c'est le cas de Théobald Smith, Francis Macfarlane Burnet, René Dubos (que citent précisément Casanova et Abel), Frank Fenner ou encore Joshua Ledenberg (Anderson, 2004 ; Méthot, 2012b). Cependant, comme Anderson le remarque, il existe tout d'abord de vraies différences conceptuelles dans la façon dont ces auteurs définissent les notions d'écologie et d'environnement. Ainsi, René Dubos définit le concept d'environnement au sens le plus large du terme, celui-ci prenant aussi bien en compte les paramètres physiques du milieu (climat, fertilité, etc.) que les facteurs de compétition entre pathogènes (en particulier dans les infections mixtes) et que les facteurs sociaux (mondialisation, accélération des flux de personnes et transports de masse, pauvreté, etc.), tandis que Smith défend une définition de l'environnement plus biologique. Par ailleurs, ces chercheurs se perçoivent tous comme dépositaires d'un discours à contre-courant du discours scientifique majoritaire dans l'explication des maladies infectieuses et tous viennent d'horizons disciplinaires relativement différents (vétérinaires, médecins, immunologistes, épidémiologistes, biologie des plantes et des animaux, etc.), utilisant des méthodes scientifiques variées (modèles mathématiques, modèles *in vivo* de simulation, etc.) – une diversité qui caractérise l'interdisciplinarité de l'écologie des maladies infectieuses. On peut expliquer le faible intérêt porté à l'écologie des maladies infectieuses en partie par les succès des autres approches (microbienne, immunologique) dans le contrôle des maladies infectieuses, en particulier avec le développement des antibiotiques et des vaccins, des thérapeutiques qui s'appuient essentiellement sur une vision centrée sur le pathogène (et plus précisément sur la génétique du pathogène incriminé) ou sur une vision centrée sur l'hôte et qui a soulevé beaucoup d'espoir quant à la possible éradication des maladies infectieuses.

Ce n'est que dans les années 1980, lorsque la communauté internationale prend conscience de la gravité de l'épidémie du SIDA d'une part et de la relative fréquence des maladies émergentes et ré-émergentes (Cohen, 2000 ; Grmek, 1993) que l'intérêt pour cette discipline va se renouveler : comprendre comment de nouveaux pathogènes émergent, ou comment des espèces non pathogènes le deviennent, suppose de s'intéresser aux interactions entre l'hôte et le pathogène, en utilisant « une métaphore plus écologiquement informée » (Lederberg, 2000) que celle de la « chasse aux microbes » (Anderson, 2004, 200) qui a prévalu pendant les développements de la microbiologie.

On peut donc tenter de dégager quelques caractéristiques communes non pas à la théorie écologique des maladies infectieuses mais aux visions écologiques de la maladie qui se sont ponctuellement développées tout au long du vingtième siècle. L'écologie des maladies infectieuses ne s'intéresse ni aux pathogènes, ni au système immunitaire de leur hôte, elle s'intéresse aux interactions hôte-pathogène. Elle implique fréquemment des modélisations mathématiques qui cherchent à prédire les « points chaud des maladies infectieuses » (c'est-à-dire les situations écologiques les plus susceptibles de provoquer une maladie émergente ou ré-émergente) ou à modéliser le mode de transmission du pathogène, ce qui suppose de bien connaître à la fois les écosystèmes du pathogène (relation du pathogène avec d'autres pathogènes ou avec d'autres mécanismes, facteurs physiques et sociaux), et de comprendre l'histoire des rapports entre l'hôte et le pathogène (mutualisme, commensalisme, parasitisme, etc.) comme le résultat des interactions dynamiques entre des organismes (l'hôte et le pathogène), qui intègre à la fois la théorie de l'évolution et une vision plus large de ce qu'est une maladie infectieuse. Il est intéressant de noter que dans le style écologique, les concepts de virulence et de pathogénicité sont définis différemment que dans le style moléculaire (Casadevall et Pirofski, 1999 ; Méthot, 2012a) : les propriétés de virulence et de pathogénicité ne sont ainsi plus caractérisées comme des propriétés intrinsèques du micro-organisme pathogène, mais comme des propriétés relatives dépendant des interactions entre le micro-organisme et l'hôte dans un environnement donné. Cette redéfinition des concepts de virulence et de pathogénicité a pour conséquence de relativiser la distinction entre les micro-organismes pathogènes et non-pathogènes.

Du point de vue de la variabilité clinique interindividuelle aux infections, la « théorie écologique » des maladies infectieuses permet donc de comprendre comment un même micro-organisme peut être pathogène dans un certain contexte environnemental et commensal dans un autre et comment un même micro-organisme n'aura pas la même virulence en fonction de l'environnement dans lequel il est et en fonction de la compétence immunologique de l'hôte.

Ainsi, si la variabilité clinique interindividuelle qui est la majeure difficulté que rencontre la théorie des germes n'est qu'imparfaitement expliquée par la théorie microbienne, la théorie immunologique et la théorie écologique (voir Tableau 6 pour une vision d'ensemble), Casanova et Abel affirment que la théorie génétique des maladies infectieuses va permettre de combler ce vide explicatif. Mais comment est née l'hypothèse selon laquelle la génétique de l'hôte pourrait combler ce manque explicatif ?

	Théorie des germes	Théorie microbienne *sensu* Abel et Casanova	Théorie immunologique *sensu* Abel et Casanova	Théorie écologique *sensu* Abel et Casanova
Définition	Spécificité un germe – une maladie + Postulats de Koch	Théorie des germes + génétique du pathogène + postulats moléculaires de Koch	L'immunité acquise contribue à la VCI*	La variabilité environnementale et les interactions hôte-pathogène contribuent à VCI*
Objet principal d'étude	Micro-organisme pathogène	Propriétés spécifiques du micro-organisme pathogène	Système immunitaire inné et acquis de l'hôte	Interactions hôte-pathogène et environnement au sens large
Figures-clés	Koch Pasteur	Koch Pasteur Falkow	Ebehrt – Metchnikoff – Burnet – Pradeu	Smith Burnet Dubos
Classes de faits de la VCI* expliquées	Aucun	Des micro-organismes d'une même espèce mais de souche différente n'ont pas le même pouvoir pathogène	Efficacité de la sérothérapie et de la vaccination dans la prévention des infections primaires et sur les infections secondaires	La virulence et la pathogénicité d'un microbe dépendent des conditions environnementales de sa rencontre avec l'hôte
Observations discordantes	Porteurs sains Résistance aux infections Variabilité des formes cliniques Variabilité de la sévérité clinique	Variabilité clinique lorsque deux individus sont infectés par la même souche du même pathogène ?	Échec des vaccins (BCGéites) ? Infections létales chez des enfants immuno - compétents ?	Variabilité clinique lorsque deux patients sont infectés par la même souche du même pathogène dans un même environnement ?

Tableau 6 : La variabilité clinique inter-individuelle est imparfaitement expliquée par la théorie des germes, la théorie microbienne, la théorie immunologique et la théorie écologique des maladies infectieuses. *VCI : variabilité clinique interindividuelle

5.1.3. De l'hypothèse génétique à la théorie génétique des maladies infectieuses

L'hypothèse génétique a une longue histoire[45]. Ainsi, la tuberculose a longtemps été considérée comme une maladie héréditaire avant l'avènement de la théorie des germes. Une fois que la théorie des germes et la théorie immunologique ont été validées par l'efficacité de la vaccination, le problème des porteurs asymptomatiques (aussi appelés « infections latentes » par Charles Nicolle) demeurait inexplicable. En Angleterre, pour expliquer ce problème, un épidémiologiste comme Karl Pearson (Pearson, 1912) et un clinicien comme Archibald Garrod (Garrod, 1931) vont proposer l'hypothèse selon laquelle le patrimoine héréditaire de l'hôte détermine la prédisposition ou la résistance aux maladies infectieuses. Par ailleurs, des études de jumeaux sur la tuberculose (Puffer, 1944), mais aussi différentes expériences sur des modèles de souris et en biologie des plantes, ont semblé renforcer cette hypothèse. Mais la découverte des antibiotiques dans les années 1940 amenuise l'intérêt que généticiens (cliniciens ou épidémiologistes) portent à la question, tant les antibiotiques semblent consacrer l'approche de la théorie des germes.

Dans les années 1950, les deux champs de l'épidémiologie génétique et de la génétique clinique des maladies infectieuses ont simultanément fait d'immenses progrès, tout en restant distincts l'un de l'autre. C'est ainsi que du côté de l'épidémiologie génétique, confirmant une prédiction faite par Haldane, Allison démontre en 1954 que les porteurs hétérozygotes de l'anémie falciforme présentent une résistance relative au paludisme à *Plasmodium falciparum*. Du côté de la génétique clinique, la mise en évidence de l'agammaglobulinémie de Bruton en 1952 (Bruton, 1952) signe la découverte des immunodéficiences primaires qui sont définies comme des maladies rares, mendéliennes, à pénétrance complète et qui sont caractérisées par des anomalies immunologiques détectables entrainant une susceptibilité à des infections multiples et récurrentes.

Entre les années 1950 et les années 1990, les deux champs sont demeurés distincts : le champ de la génétique clinique étudie les prédispositions mendéliennes à des infections multiples faisant triompher l'hypothèse *« rare variants, rare diseases »,*

[45] Les limites de ce travail ne nous permettent pas de proposer une historiographie critique des origines de cette hypothèse génétique. Nous nous contentons donc de reprendre ici les principaux repères donnés par Casanova et Abel dans leur article de référence de 2013 (Casanova et Abel, 2013a).

(un gène - plusieurs infections - phénotype rare), tandis le champ de l'épidémiologie génétique étudie les prédispositions polygéniques faisant triompher l'hypothèse *« common variants, common diseases »* que nous avons déjà évoquée à plusieurs reprises dans le chapitre 3 (plusieurs gènes ou variants alléliques - une infection - phénotype commun) :

> « Jusqu'au milieu des années 1990, le paradigme dominant dans le champ de la génétique des populations étaient « associés à » de multiples variants communs, ce qu'illustre bien le cas du trait drépanocytaire qui confère une résistance au paludisme sévère (plusieurs gènes, une maladie). À l'inverse, le champ de la génétique clinique était dominé par l'idée que les traits mendéliens rares confèrent une prédisposition à des infections multiples et récurrentes (un gène, plusieurs infections). Les deux champs se sont ignorés mutuellement jusqu'à ce que l'on démontre l'effet des défauts monogéniques chez les individus et des gènes majeurs chez les populations (un gène, une maladie).» (Casanova et Abel, 2013, p. 225)

C'est donc la découverte dans les années 1990 de deux nouveaux mécanismes (le concept de prédisposition mendélienne à une seule infection et le concept de gène majeur) qui va permettre aux deux champs de la génétique clinique et de l'épidémiologie clinique de communiquer, ce qu'illustre parfaitement l'alliance de Laurent Abel, issu de l'épidémiologie génétique et déjà célèbre pour ses travaux sur le concept de gène majeur dans la lèpre (Abel et Demenais, 1988 ; Abel *et al.,* 1989*, 1995* ; Abel et Dessein, 1997), et de Jean-Louis Casanova, généticien et pédiatre, qui a découvert en 1996 les premiers exemples de prédisposition mendélienne aux infections mycobactériennes (Altare *et al*., 1998 ; Casanova, 2000 ; Casanova *et al.*, 1996). Ils fondent ensemble en 2001 le Laboratoire de Génétique des Maladies Infectieuses de l'Hôpital Necker et commencent alors à développer leur concept de théorie génétique des maladies infectieuses sur une centaine d'articles en défendant l'idée que ces deux nouveaux mécanismes (gène majeur d'une part et prédisposition mendélienne à une seule infection d'autre part) viennent « combler la distance » entre le champ de la génétique clinique et celui de la génétique des populations (Abel et Casanova, 2000).

Si elle est défendue essentiellement par cette équipe, notons que la théorie génétique des maladies infectieuses a acquis une solide renommée : les équipes de Casanova et Abel publient ainsi régulièrement dans des journaux scientifiques de rang A tels que *Science*, *Nature*, the *American Journal of Human Genetics*, et ont coécrit le chapitre sur la génétique des maladies infectieuses dans l'ouvrage de génétique humaine de référence, édité par Vogel et Motulsky (Alcaïs, Abel, *et al.,* 2010). Le Laboratoire de Génétique des Maladies Infectieuses s'est d'ailleurs étendu avec le développement d'une branche américaine au *Rockefeller Institute*, dirigée par Jean-Laurent Casanova et avec des collaborations très larges à travers le monde (avec l'équipe de Luis-Quintano Murci en génétique des populations à l'Institut Pasteur en France, l'équipe d'Erwin Schurr, directeur du Centre de recherche de McGill au Canada sur la résistance de l'hôte à la maladie, ou l'équipe de Jamila El Baghdadi, professeur de génétique à l'hôpital de Rabat, au Maroc).

5.2. Présentation de la théorie génétique des maladies infectieuses

À présent que nous avons brièvement exposé le cadre théorique et les raisons historiques pour lesquelles Abel et Casanova ont développé la théorie génétique des maladies infectieuses, il nous faut exposer en détail les quatre mécanismes génétiques fcommuns précédemment évoqués et en quoi ces mécanismes génétiques forment une théorie génétique des maladies infectieuses.

5.2.1. Quatre mécanismes génétiques communs non mutuellement exclusifs

La théorie génétique des maladies infectieuses est fondée sur l'identification de quatre mécanismes génétiques communs à toutes les maladies infectieuses qui sont résumés dans la Table 2. Deux de ces mécanismes (la prédisposition mendélienne à plusieurs infections et la prédisposition polygénique à une ou plusieurs infections) sont connus ou soupçonnés depuis les années 1950. Les deux autres mécanismes ont été décrits beaucoup plus récemment (Casanova et Abel, 2013): il s'agit de la prédisposition mendélienne à une infection et du concept de gène majeur / résistance à une infection.

5.2.1.1. La prédisposition mendélienne à plusieurs infections et les immunodéficiences primaires conventionnelles

Dans ce premier mécanisme, une ou plusieurs mutations dans un gène, dont les produits sont impliqués dans la réponse des cellules de l'immunité humorale ou cellulaire, entraînent une susceptibilité à de multiples infections. Le premier[46] exemple de ce type de prédisposition est l'agammaglobulinémie liée à l'X, découverte par Ogden Bruton en 1952, en examinant un jeune garçon de huit ans ayant présenté quatorze épisodes d'infections à pneumocoque. L'analyse du sérum révélait une absence d'immunoglobulines (Bruton, 1952 ; Bruton *et al.,* 1952), mais l'administration régulière d'immunoglobulines IgG a permis de sauver le jeune patient. L'agammaglobulinémie liée à l'X est causée par des mutations dans le gène de la tyrosine kinase de Bruton. La tyrosine kinase de Bruton joue un rôle essentiel dans la maturation des lymphocytes B dans la moelle osseuse. Lorsque ce gène est muté, les lymphocytes B immatures ne peuvent se développer en des lymphocytes B fonctionnels, entraînant une susceptibilité à des infections bactériennes multiples dans l'enfance des individus (principalement masculins[47]) affectés. D'autres exemples célèbres de déficits immunitaires primaires sont la neutropénie congénitale sévère ou l'immunodéficience sévère combinée. Plus de deux cent maladies de ce type ont été découvertes (Al-Herz *et al.,* 2011) qui sont également appelées « immunodéficiences primaires conventionnelles » (« *Conventional Primary immunodeficiencies* », « *Conventional PID* ») ou désordres immunitaires primaires (DIP), par opposition aux déficits immunitaires acquis comme celui qui résulte d'une infection au Virus de l'Immunodéficience Humaine (VIH) et par opposition aux « nouveaux déficits

[46] Pour être plus précis, l'agammaglobulinémie est le premier déficit immunitaire primaire à avoir été reconnu comme tel au moment de sa description. En effet, comme le font remarquer Casanova et Abel, les déficits immunitaires primaires ont été longtemps « définis par l'existence d'un phénotype « immunologique », au sens hématopoïétique du terme, c'est-à-dire par une anomalie moléculaire ou cellulaire détectable dans le sang » (Casanova et Abel, 2007b, p. 617). Ainsi, l'épidermodysplasie verruciforme, également appelée syndrome de Lutz-Lewandowsky, est une dermatose caractérisée une infection chronique au papillomavirus humain (HPV) entraînant l'apparition de lésions cutanées polymorphes et présentant un risque élevé de cancer de la peau. L'épidermodysplasie verruciforme a été découverte en 1946, mais en l'absence de tout phénotype immunologique clair, elle n'a été reconnue comme un déficit immunitaire primaire qu'au moment où le rôle des gènes EVER-1 et EVER-2 a été mis en évidence dans sa physiopathologie (Ramoz *et al.*, 2002).

[47] Seuls les individus masculins sont affectés, puisqu'il s'agit d'une prédisposition liée au chromosome X. Chez les individus de sexe féminin, l'existence d'un second chromosome X permet de pallier cette déficience.

immunitaires primaires » qui sont liés au second mécanisme de la théorie génétique des maladies infectieuses que nous décrivons ci-après.

Quatre caractéristiques peuvent donc être retenues pour définir ce mécanisme : (1) La prédisposition est dite « mendélienne » au sens large du terme : une ou plusieurs mutations, héritée(s), récessive(s) ou dominante(s) sur un seul gène qui peut être situé sur les chromosomes autosomiques ou sexuels (X ou Y), sont impliqués dans l'apparition de la maladie. Notons que la plupart des immunodéficiences primaires autosomiques sont récessives du fait de leur extrême gravité : les formes autosomiques dominantes sont rarement compatibles avec la survie (Casanova *et al., 2005* ; Lawrence *et al., 2005)*. (2) Ces mutations entrainent un déficit immunitaire, dans l'immunité cellulaire ou dans l'immunité humorale. (3) Ce déficit n'est pas spécifique d'une seule infection, mais touche une ou plusieurs des lignées hématopoïétiques du système immunitaire qui sont impliquées dans une réponse peu spécifique du système immunitaire aux infections, entrainant une susceptibilité à une classe d'infections (les infections bactériennes par exemple dans le cas de l'agammaglobulinémie liée à l'X). (4) Ces déficits peuvent être corrigés par des thérapies de substitution : injection de gammaglobulines dans le cas de l'agammaglobulinémie liée à l'X, mais aussi greffe de moelle osseuse ou thérapie génique dans l'immunodéficience sévère combinée, injection de facteurs de croissance hématopoïétiques comme le G-CSF, dans le cas de la neutropénie congénitale sévère. (5) Ces maladies sont rares, chroniques et se manifestent tôt dans l'enfance.

5.2.1.2. La prédisposition mendélienne à une infection ou les nouvelles immunodéficiences primaires

Dans ce second mécanisme, une ou plusieurs mutations dans un gène causent une susceptibilité à une seule infection. Prenons l'exemple de l'encéphalite herpétique. L'herpès virus de type 1 (HSV-1) infecte environ 80% de la population mondiale, et a la particularité après la primo-infection de rester à l'état de latence dans les cellules de l'hôte infecté, dont il ne sort qu'à l'occasion de réactivations ou récurrences. L'infection à HSV-1 est la plupart du temps soit complètement asymptomatique, soit responsable de manifestations cutanéo-muqueuses orales

bénignes (herpès labial, gingivo-stomatite aiguë). Cependant, de rares individus infectés par HSV-1 vont développer une méningo-encéphalite herpétique aux manifestations cliniques extrêmement bruyantes (fièvre, syndrome confusionnel, troubles de la conscience, crises convulsives, déficits moteurs ou cognitifs). Non traitée, la méningo-encéphalite herpétique est mortelle dans 70% des cas. La mortalité tombe à 20% chez les patients traités mais d'importantes séquelles neurocognitives persistent à vie chez plus de la moitié des patients (Whitley, 2006). L'encéphalite herpétique, qui peut survenir soit au moment de la primo-infection, soit au moment de réactivation du virus lors de récurrences, demeure cependant la forme la plus commune d'encéphalite sporadique dans les pays occidentaux (1 cas pour 250 000/an), affectant en priorité les enfants de moins de trois ans et les adultes de plus de 50 ans. Comment comprendre que seule une très petite fraction d'individus infectés par HSV1 développent l'encéphalite herpétique ? En 2006, une première prédisposition mendélienne (autosomique récessive) à l'encéphalite herpétique, touchant le gène UNC-93B1 a été mise en évidence chez deux enfants ayant présenté un épisode d'encéphalite herpétique (Casrouge *et al., 2006)*. La protéine UNC-93B1 interagit avec la voie des récepteurs « Toll-Like », des protéines membranaires qui reconnaissent des structures moléculaires communes à de nombreux micro-organismes pathogènes et qui jouent un rôle crucial dans le déclenchement de la réponse immunitaire innée. En l'occurrence, la mutation du gène UNC-93B1 inactive complètement la protéine UNC-93B1, ce qui empêche la reconnaissance des ARN intermédiaires de HSV-1 par les récepteurs Toll-Like du système nerveux central. Cette absence de reconnaissance du virus compromet la production d'interféron qui joue un rôle essentiel dans l'immunité innée à HSV-1 : le virus peut se multiplier dans le système nerveux central et entraine la mort des cellules nerveuses infectées. Plusieurs mutations mendéliennes dans d'autres gènes dont les produits interagissent avec la voie des récepteurs Toll-Like ont été mises en évidence depuis : en particulier la mutation autosomique dominante du gène HSE2 qui code directement pour le récepteur Toll-Like 3 ou TLR3 (Zhang *et al.,* 2007), une mutation autosomique récessive aboutissant à un défaut complet du TLR3 (Guo *et al.,* 2011), plusieurs mutations hétérogènes du gène TBK1 qui interagit avec TLR3 (Herman *et al.,* 2012, p. 1) et plusieurs mutations autosomiques dominantes et récessives du gène TRIF

(Sancho-Shimizu *et al.,* 2011) qui interagit avec TLR4 (un autre récepteur de la voie des Toll-Like qui est impliqué dans la production d'interféron en réponse à HSV-1). Ces différentes découvertes permettent de redéfinir l'encéphalite herpétique comme une collection de prédispositions mendéliennes monogéniques.

Les maladies infectieuses pour lesquelles ont été mises en évidence ce mécanisme de prédisposition mendélienne un gène-une infection sont appelées les « nouvelles immunodéficiences primaires » (« *Novel primary immunodeficiencies* »), étant donné qu'elles ont été découvertes récemment, bien après les immunodéficiences primaires conventionnelles qui sont des prédispositions mendéliennes à un spectre beaucoup plus large d'infections. Selon Casanova et Abel, il y a deux explications complémentaires à cette découverte tardive. La première explication est que les immunodéficiences primaires ont été d'abord définies en fonction d'un phénotype clinique bruyant (de multiples infections récurrentes) et d'anomalies immunologiques détectables (absence d'une lignée hématopoïétique, etc.) : aucune attention n'a donc été portée aux phénotypes simples mais sévères, c'est-à-dire aux infections sévères isolées chez des individus considérés par ailleurs comme immunocompétents. D'autre part, si l'usage des antibiotiques et la généralisation de la vaccination ont permis la découverte des immunodéficiences primaires en permettant aux enfants qui en étaient atteints de survivre assez longtemps pour multiplier les infections et attirer l'attention des médecins sur ces phénotypes étonnants, c'est justement l'utilisation des antibiotiques et la généralisation de la vaccination qui ont masqué le caractère atypique d'infections sévères isolées, c'est-à-dire l'existence des nouvelles immunodéficiences primaires :

> « Les immunodéficiences primaires ont été alors définies comme des maladies rares, mendéliennes, à pénétrance complète, à début précoce avec de multiples infections récurrentes et opportunistes et des anomalies immunologiques manifestes (ou au moins détectables). […] Avec le recul, cette définition étroite des études sur les immunodéficiences primaires fondées sur des infections récurrentes détectables seulement après l'utilisation des antibiotiques dans les populations humaines a inévitablement entrainé une définition étroite des individus immunodéficients (que cette déficience soit héritée ou acquise) et peut être considérée comme un biais de détermination.

> Bien que fructueuse, cette erreur conceptuelle de départ explique l'oxymore bien trop communément utilisé et au final erroné que l'on trouve dans les études de cas en médecine à propos « d'infection fatale ou potentiellement mortelle chez un individu immunocompétent ». » (Casanova et Abel, 2013, p. 224)

Ces nouvelles immunodéficiences primaires peuvent donc être caractérisées comme : (1) des prédispositions mendéliennes au sens large (2) qui entrainent souvent un déficit de l'immunité innée plutôt qu'un déficit de l'immunité adaptative, (3) qui concernent une seule infection plutôt commune (ou une seule famille de micro-organismes) (4) mais qui produisent une infection sévère et la plupart du temps isolée (les récurrences sont très rares, car l'immunité adaptative non touchée par le déficit se développe alors, créant une protection relative aux récurrences), (5) dont l'âge de manifestation dépend bien sûr de l'âge à laquelle l'individu atteint est mis en contact avec le micro-organisme, mais qui intervient en moyenne plutôt à l'âge adulte que dans l'enfance.

La découverte de ces nouvelles immunodéficiences primaires a un impact théorique conséquent. (1) Elle entraine une redéfinition des immunodéficiences primaires et un élargissement de leurs classifications. (2) Plus généralement, elle suppose une redéfinition du concept d'immunocompétence – l'immunodéficience est ainsi vue non pas comme un phénomène rare, mais comme un phénomène majeur affectant probablement chaque individu à des degrés variés et qui constituent « la règle plutôt que l'exception » (Casanova et Abel, 2004, 2005):

> « L'immunocompétence de l'hôte pourrait être définie de façon étroite comme la capacité à gérer une infection microbienne sans développer de symptômes. Dans un sens plus large, elle pourrait être vue comme la survie à une infection, indépendamment des traitements préventifs (par exemple, les vaccins) ou curatifs (par exemple, les antibiotiques). Nous préférons définir **l'immunocompétence comme le statut immunologique qui résulte d'une infection qui se guérit seule – c'est-à-dire qui a une issue favorable, avec ou sans symptôme, et sans aucun traitement spécifique**. C'est un concept opérationnel qui peut être appliqué à n'importe quel hôte infecté à n'importe quel âge par n'importe quelle charge de n'importe quel microbe. Bien que la

> plupart des humeurs puissent contrôler la plupart des infections, **une seule infection potentiellement mortelle est suffisante pour définir une immunodéficience**. […] L'espérance de vie humaine n'a augmenté substantiellement que dans une petite partie riche du monde, et cette augmentation doit être attribuée à des améliorations des conditions d'hygiène (qui diminuent l'exposition aux microbes), aux vaccins (qui diminuent les maladies cliniques) et aux médicaments antimicrobiens (qui diminuent les issues fatales), – non pas à une amélioration intrinsèque, évolutive de l'immunité. L'immunité naturelle garantit la survie de l'espèce, plutôt que celle de chaque individu. […] Par contraste avec ce que l'on pourrait imaginer, en ce qui concerne l'exposition au monde microbien, **l'immunocompétence humaine (à tous les micro-organismes) est rare et l'immunodéficience (à au moins un micro-organisme) est la règle plutôt que l'exception**[48]». (Casanova et Abel, 2004, p. 56)

5.2.1.3. Le concept de gène majeur / de résistance à une infection

Le concept de « gène majeur » ou de « locus majeur » a été développé dès les années 1960 (Edwards, 1969 ; Lalouel *et al., 1983)*, dans le contexte de l'analyse de ségrégation complexe, une méthode statistique fondée sur un modèle d'héritabilité dans lequel un phénotype donné peut résulter de l'action conjointe d'un gène majeur, d'une composante polygénique et de facteurs environnementaux, afin de comprendre le phénomène de la pénétrance incomplète. Comme nous l'avons mentionné précédemment, la pénétrance est la fréquence d'individus qui portent un allèle particulier et qui expriment aussi le trait associé à cet allèle. Pour une mutation associée à une maladie donnée, la pénétrance peut être incomplète, ce qui signifie que seule une partie des individus porteurs de l'allèle donné manifesteront des symptômes de la maladie correspondante. Un gène majeur est responsable d'une déficience immunitaire mais sa pénétrance peut être diminuée sous l'effet combiné d'autres gènes et de l'environnement. La notion centrale du concept de gène majeur est que seul un gène muté cause la maladie correspondante, mais que d'autres gènes ou facteurs environnementaux sont susceptibles d'influencer l'expression de ce gène,

[48] C'est nous qui soulignons.

expliquant ainsi sa pénétrance variable. D'après Casanova, le premier exemple de gène majeur est la découverte de l'influence du gène HbS sur la drépanocytose :

> « Le trait HbS peut être considéré comme le premier gène majeur identifié dans une maladie infectieuse commune, en raison à la fois de sa fréquence dans certaines populations africaines et de son effet estimé sur les formes sévères de paludisme. L'effet d'un gène majeur diffère d'un effet mendélien en ce sens qu'il démontre une pénétrance basse (c'est-à-dire, un effet incomplet) due à une plus grande influence à la fois de l'environnement et d'autres gènes chez des individus génétiquement prédisposés. Cependant, de tels effets peuvent être tout de même considérables au niveau individuel, et une méta-analyse récente et large a estimé que le risque relatif de développer une forme sévère de paludisme pour les hétérozygotes HbAS était d'approximativement 0.09% (avec une intervalle de confiance à 95% de 0.06 à 0.12), par rapport aux homozygotes HbAA (c'est-à-dire une baisse du risque relatif d'un ordre de grandeur). Le variant HbS semble également être le déterminant génétique le plus fort de la susceptibilité / résistance aux formes sévères de paludisme dans deux associations pangénomiques récentes. Un autre variant de l'hémoglobine, HbC, augmente également la résistance à la malaria à *Plasmodium falciparum*, cet effet étant le plus fort chez les homozygotes HbCC (qui ont également une forme modérée d'anémie hémolytique) avec un risque relatif de 0.27 (intervalle de confiance à 95% de 0.11 à 0.63). Cette résistance récessive pourrait rendre compte de la distribution spécifique du variant HbC en Afrique de l'Ouest et pour le petit nombre de mutations qui ont été reportés pour HbS. » (Casanova et Abel, 2013, p. 226–227)

Ainsi, le concept de « résistance » est le miroir du concept de gène majeur : certaines mutations spécifiques sur un allèle donné confèrent une résistance à un pathogène donné parce qu'elles empêchent l'expression des récepteurs dont le micro-organisme a besoin pour pénétrer dans les cellules. C'est le cas pour le paludisme causé par *Plasmodium vivax*. *Plasmodium vivax* est l'un des quatre parasites responsables du paludisme. Pour pénétrer dans les globules rouges, *Plasmodium vivax* doit se fixer sur un co-récepteur des antigènes Duffy, appelé DARC et situé à la surface des érythrocytes. Une mutation d'un seul nucléotide sur le promoteur du gène DARC

empêche l'expression du récepteur DARC à la surface des érythrocytes, conférant aux porteurs de cette mutation une résistance au paludisme causé par *Plasmodium Vivax* (Miller et al., 1976). Un autre cas de résistance célèbre est l'exemple de la résistance au virus de l'immunodéficience humaine de type 1 (VIH-1), conférée par des mutations qui affectent le domaine extracellulaire d'un récepteur à chimiokine, CCR5. Ce récepteur est un co-récepteur de CD4, dont la liaison au VIH-1 est nécessaire pour que celui-ci puisse pénétrer dans les lymphocytes T. Les individus homozygotes pour la mutation la plus commune du gène CCR5 (une délétion sur la 32^{e} paire de bases du gène – D32), sont résistants à l'infection au virus VIH-1 (Dean *et al.,* 1996 ; Liu *et al.,* 1996).

5.2.1.4. La prédisposition polygénique à une ou plusieurs infections

L'hérédité polygénique diffère du concept de gène majeur. Elle implique que le phénotype global résulte non pas d'un seul gène influencé par d'autres gènes et par des facteurs environnementaux, mais de l'effet combiné d'un grand nombre de loci. En fonction du nombre et de l'impact relatif des gènes influençant la maladie, on peut distinguer entre une prédisposition oligogénéique et une « vraie » prédisposition polygénique. L'oligogénicité implique que le phénotype dépend de deux ou plusieurs gènes majeurs, tandis que les autres loci ont une contribution relativement faible au phénotype global. La prédisposition polygénique « vraie » n'implique aucun gène majeur et l'apparition de la maladie dépend d'un large nombre de loci, chacun d'entre eux ayant une contribution relativement faible au phénotype global. Il est essentiel de comprendre que comme le concept de gène majeur, la prédisposition polygénique est un concept dont on a souvent fait l'hypothèse en génétique des populations (en particulier dans le cadre des infections associées au complexe HLA, (Alcaïs *et al.,* 2009)), mais qu'en raison de la multiplication des variants génétiques impliqués par ce concept, il n'y a pas à proprement parler de démonstration d'une prédisposition polygénique à une ou plusieurs infections au niveau d'un individu ou même au niveau d'un groupe ethnique particulier. En effet, pour prouver l'existence d'un mécanisme de prédisposition polygénique à une ou plusieurs infections chez un individu, il faudrait pouvoir démontrer l'effet cumulatif de très nombreux variants alléliques à effet faible

(et donc très difficilement détectables dans les associations pangénomiques ou dans les analyses de ségrégation) et l'influence de l'environnement sur ces effets chez un même individu ou dans une même population :

> « L'identification d'une prédisposition véritablement polygénique nécessite un large nombre d'individus, à la fois en raison du faible effet attendu attribuable à chaque gène et en raison de la nature additive de ces effets génétiques. C'est pourquoi il n'est pas surprenant que la preuve de principe de tels mécanismes génétiques à la fois au niveau de la population et au niveau de l'individu n'ait pas encore été fournie par des études humaines, mais seulement par des études de la susceptibilité des maladies infectieuses dans les modèles animaux d'infections expérimentales. […] Il a souvent été rapporté que plusieurs gènes avaient une influence potentielle sur le déclenchement d'une maladie infectieuse donnée chez les humains, mais il n'y a pas à l'heure actuelle de preuve d'une prédisposition polygénique *per se* dans des populations spécifiques et chez des individus de l'espèce humaine. Par exemple, on a rapporté qu'un certain nombre de gènes (comme le locus HLA-DR, NRAMP1 et ILI2RB1) joueraient un rôle dans la tuberculose mais on n'a jamais déterminé si ces gènes agissent *indépendamment* et *additivement* sur le *même* phénotype dans le *même*[49] échantillon. » (Alcaïs *et al.,* 2009, p. 2511)

À présent que nous avons présenté les quatre mécanismes génétiques de la théorie génétique des maladies infectieuses (Tableau 7), il nous faut discuter les conséquences de l'identification de ces quatre mécanismes communs sur l'émergence d'un nouveau concept de continuum génétique des maladies infectieuses (5.2.2) et sur l'hypothèse d'une architecture allélique dépendante de l'âge des maladies infectieuses (5.2.3).

[49] C'est nous qui soulignons.

Mécanisme	Description	Exemple
Prédisposition mendélienne à des infections multiples	Un seul gène, pénétrance complète, des infections multiples	***Agammaglobulinémie liée à l'X :*** Mutations du gène de la tyrokinase kinase Btk ⇒ lymphocytes B immatures ⇒ infections bactériennes multiples
Prédisposition mendélienne à une seule infection	Un seul gène, pénétrance complète, une seule infection	***Encéphalite simple à Herpès virus :*** Mutation autosomique récessive du gène UNC93B : ⇒ mauvaise reconnaissance d'HSV1 par le SNC ⇒ production défectueuse d'interféron ⇒réplication virale dans le SNC
Gène majeur / résistance à une seule infection	Un gène majeur, haute pénétrance, une seule infection	***Paludisme causé par P. vivax :*** Mutations dans le promoteur du gène DARC ⇒ absence de DARC, corécepteur de P. vivax ⇒ P. vivax ne peut pénétrer dans les érythrocytes d'où « résistance »
Prédisposition polygénique à une ou plusieurs infections	Gènes multiples, pénétrance basse, une ou plusieurs infection(s)	***Infections associées au complexe HLA***

Tableau 7 : Quatre mécanismes génétiques communs dans la théorie génétique des maladies infectieuses. Pour chacun de ces mécanismes, une description est proposée et un exemple de maladie est donné.

5.2.2. Des mécanismes génétiques communs au continuum génétique de la maladie infectieuse

Il semblerait logique que la différenciation de quatre mécanismes génétiques dans la théorie génétique des maladies infectieuses aboutisse à une catégorisation des maladies infectieuses en quatre classes, chacune correspondant à un type de mécanisme. Or, Casanova et Abel insistent sur le fait que ces mécanismes sont « communs » ou « transversaux » à toutes les maladies infectieuses (Alcaïs *et al.,* 2009 ; Alcaïs, Abel, *et al.,* 2010 ; Casanova et Abel, 2007a, 2013). Ces termes sont extrêmement importants car ils signifient qu'il n'y a pas d'équation une maladie – un mécanisme, mais au contraire, qu'une même maladie peut survenir par plusieurs mécanismes différents. Par exemple, l'infection à *Mycobacterium tuberculosis* ne relève pas du même mécanisme génétique dans toutes les populations : selon les populations, il y a des cas de prédisposition mendélienne à la tuberculose, des cas où

un gène majeur est impliqué et des cas de prédisposition polygénique supposée (Abel et Casanova, 2000 ; Alcaïs, Fieschi, *et al.,* 2005 ; Casanova, 2001 ; Al-Muhsen et Casanova, 2008).

Ainsi l'identification de quatre mécanismes génétiques communs n'entraîne pas une distinction absolue entre quatre catégories mutuellement exclusives de maladies infectieuses, comme on peut le voir sur la Figure 23. Au contraire, Casanova et Abel parlent d'un « spectre continu » de la génétique des maladies infectieuses, qui s'explique à la fois parce que la distinction entre les quatre mécanismes génétiques est relative et non pas absolue (ainsi, certaines maladies sont dites dans une position intermédiaire, entre une prédisposition à gène majeur et une prédisposition polygénique), mais aussi parce qu'une même maladie peut advenir par l'intermédiaire de différents mécanismes génétiques. C'est précisément parce qu'il n'y a pas de correspondance entre un mécanisme génétique et une catégorie de maladie infectieuse que l'on peut parler d'un véritable continuum mécaniste et non d'une simple typologie.

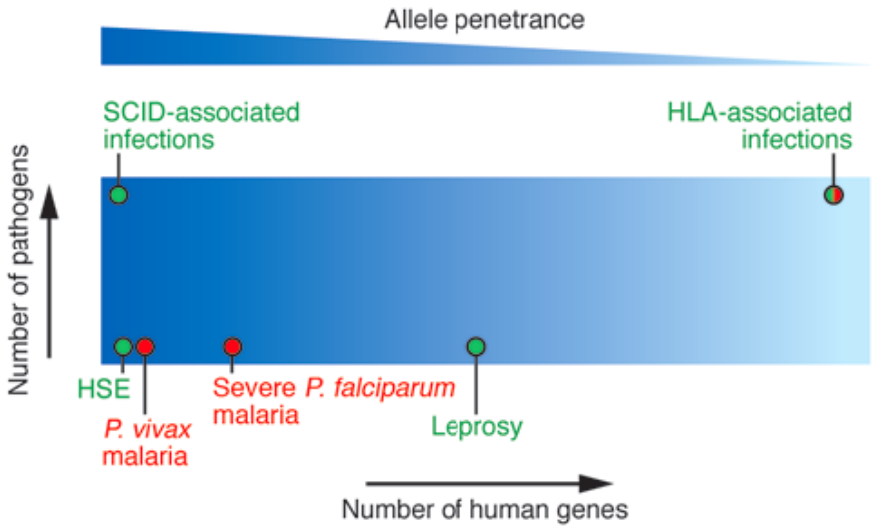


Figure 23 : Représentation schématique du continuum génétique de la maladie infectieuse (Alcaïs *et al.,* 2009, p. 2508)**.** L'ordonnée correspond au nombre d'infections à risque de développement. L'abscisse désigne le nombre de gènes qui entrent en jeu dans la maladie donnée. La pénétrance allélique est représentée par le triangle au-dessus du graphique : lorsqu'un seul gène ou un gène majeur est impliqué, la pénétrance est élevée. À l'inverse, quand de nombreux gènes sont impliqués et que chaque gène a un effet limité sur le phénotype global, la pénétrance est basse. Finalement le dégradé de bleu représente le continuum génétique des maladies infectieuses et est fortement corrélé avec la pénétrance allélique comme le montre le code couleur commun.

Le terme de continuum génétique acquiert ici une signification complètement différente de celle que nous avions étudié au chapitre 1. Ainsi, dans le manuel de génétique médicale publié par le Collège National de Enseignants et des Praticiens de Génétiques Médicales (CNEPGM) en 2004 et reproduit ci-dessous (Figure 24), le graphique, légendé « action conjointe des facteurs génétiques et environnementaux de la maladie », consiste en une seule ligne divisée en trois segments (noir, gris et blanc). Chacun de ces segments correspond à une catégorie spécifique de maladie : le segment noir représente « les maladies les plus génétiques » tandis que le segment blanc représente « les maladies les plus environnementales ». Entre ces deux extrêmes, le dégradé de gris indique des maladies qui dépendent à la fois de facteurs génétiques et de facteurs environnementaux. Autrement dit, c'est une représentation typique d'un continuum génétique compris sous l'angle du problème de la sélection causale : certaines maladies sont plus génétiques que d'autres et ce qui est en jeu, c'est de déterminer en quelles proportions respectives les facteurs génétiques et environnementaux causent la maladie. Dans ce type de continuum génétique, toutes les maladies infectieuses décrites ci-dessus se trouveraient dans la partie blanche, « la plus environnementale », du continuum.

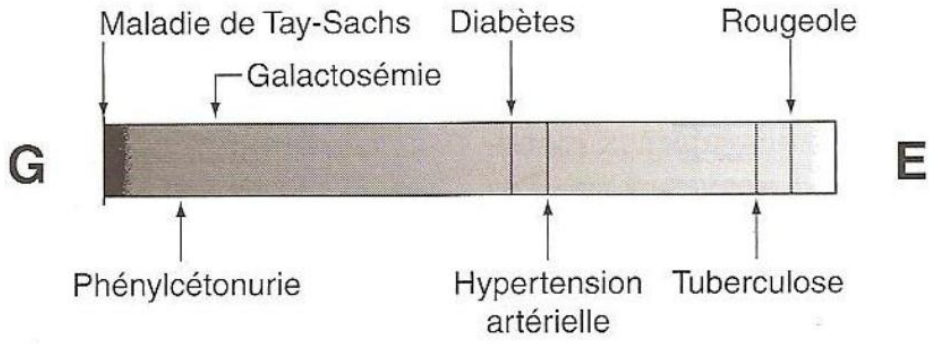


Figure 24 : Diagramme de l'action conjointe des facteurs génétiques (G) et environnementaux (E) sur les maladies (Collège national des enseignants et praticiens de génétique médicale, 2004, 21)

Le continuum génétique de la théorie génétique des maladies infectieuses diffère de la version du CNEPGM sous deux aspects. Premièrement, le problème de la sélection causale n'apparaît pas : il n'y a pas de distinction entre les maladies fondées sur le degré auquel les gènes et l'environnement influencent les phénotypes. En effet,

le dégradé continu de bleu représente la pénétrance allélique, c'est-à-dire, la fraction des individus qui ont le gène et la maladie correspondante. Mais une maladie avec une pénétrance basse n'est pas une maladie avec une influence génétique moindre. L'hérédité polygénique ne pose pas moins d'influence génétique que la prédisposition mendélienne, en revanche il y a une différence dans la façon dont les gènes causent la maladie selon les mécanismes. Deuxièmement, alors qu'il est impossible de distinguer les différents mécanismes causaux à l'œuvre dans une même maladie sur la représentation traditionnel du continuum génétique du CNEPGM, ceci est possible dans les représentations utilisées par Abel et Casanova. Par exemple, la tuberculose devrait apparaître à trois points différents de la Figure 23, puisqu'elle peut être causée par au moins trois mécanismes génétiques différents, comme nous l'avons mentionné plus haut.

Quelles sont les conséquences épistémologiques de ce continuum mécaniste ? D'abord, il fournit une explication unifiée de certaines classes de faits de la variabilité clinique interindividuelle aux infections du point de vue génétique. Pour chaque maladie infectieuse, on suppose qu'un ou plusieurs de ces mécanismes peuvent être instanciés pour expliquer pourquoi une fraction des individus infectés manifeste des symptômes tandis que la majorité de la population demeure asymptomatique. Ainsi, la théorie génétique des maladies infectieuses représente un gain important dans notre compréhension de la pathogénèse des maladies infectieuses, comparée aux autres explications fournies par la théorie des germes, la théorie microbienne, la théorie immunologique et la théorie écologique qui ne pouvaient pas rendre compte du problème des porteurs asymptomatiques ou de la variété des formes cliniques manifestées par les individus atteints d'une même maladie. Surtout et c'est en cela qu'elle est une explication large de la variabilité interindividuelle, la théorie génétique des maladies infectieuses fournit une explication qui est satisfaisante au niveau des individus comme au niveau de la population. Ainsi, la théorie des germes ne pouvait fournir une explication de la tuberculose que pour les individus qui avaient des symptômes cliniques de la tuberculose (soit environ 10% d'une population infectée). Au contraire, la théorie génétique des maladies infectieuses permet deux types d'explications. Au niveau de la population, elle permet de rendre compte de la pluralité des mécanismes génétiques impliqués dans une même maladie

(éventuellement un mécanisme donné peut correspondre à une sous-population de la population infectée). Mais au niveau individuel, il est possible de distinguer entre ces différents mécanismes pour expliquer pourquoi tel individu, dans ce cas particulier, a développé la tuberculose.

5.2.3. L'âge de début de la maladie, une variable prédictive de l'architecture génétique des maladies infectieuses

Pour compléter cette notion de continuum génétique des maladies infectieuses, Casanova et Abel ont fait l'hypothèse qu'on pouvait prédire l'architecture génétique d'une maladie en fonction de l'âge, ou plus exactement qu'on pouvait prédire quel mécanisme génétique était impliqué en fonction des courbes de mortalité d'une maladie infectieuse donnée. Casanova et Abel ont développé l'hypothèse selon laquelle les formes très sévères de l'infection de l'enfant révélaient la plupart du temps des prédispositions mendéliennes à une ou plusieurs infections (mécanismes 1 ou 2 dans la Tableau 7), tandis que les infections secondaires ou les réactivations chez l'adulte jeune révèlent l'influence d'un gène majeur et qu'enfin, plus l'âge de déclenchement de la maladie est tardif, plus les infections secondaires ou les réactivations sévères révèlent plutôt une prédisposition polygénique aux infections. Cette hypothèse, représentée dans la Figure 25, est forte car elle permet à la fois d'expliquer la sévérité, voire le caractère fatal d'un grand nombre d'infections infantiles (par rapport au caractère relativement bénin de ces mêmes infections dans leur forme adulte) et de doter la théorie génétique des maladies infectieuses d'un outil de prédiction qui est aussi un outil de confirmation ou d'infirmation de la théorie génétique des maladies infectieuses.

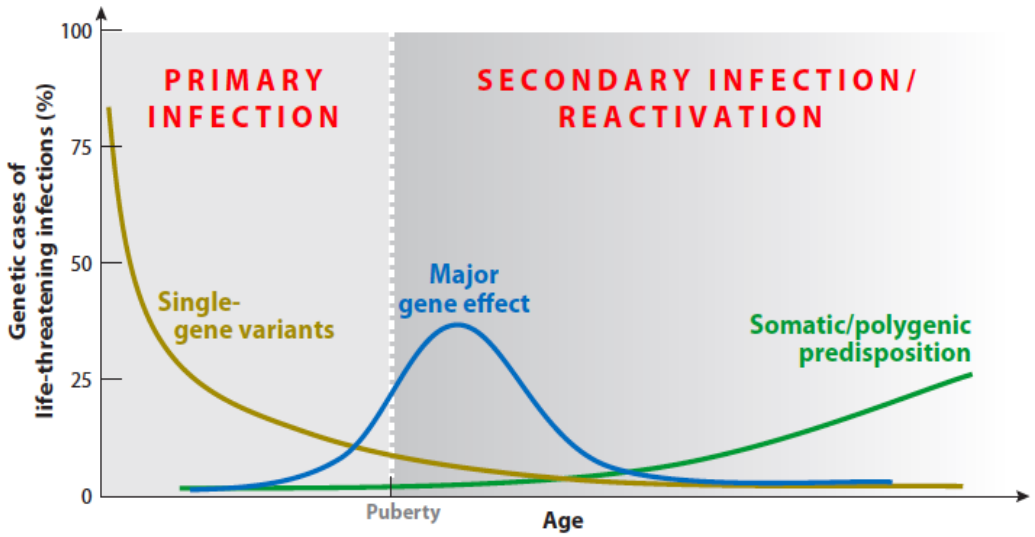


Figure 25 : L'hypothèse d'une architecture génétique des maladies infectieuses (Casanova et Abel, 2013, p. 226). Les infections primaires mortelles ou potentiellement fatales de l'enfance seraient déterminées par des mécanismes mendéliens ou de gène majeur, tandis que les infections moins sévères des adultes (réactivations symptomatiques ou infections secondaires) seraient déterminées par des mécanismes polygéniques ou somatiques.

Le point de départ de cette hypothèse est l'observation des courbes de mortalité des maladies infectieuses au fil des siècles. On peut ainsi constater qu'en dépit de la réduction relative de la mortalité de certaines infections, les courbes de mortalité ont toujours la même forme en U ou en L (voir Figure 26), c'est-à-dire que pour une même maladie infectieuse, la mortalité est plus forte aux âges extrêmes de la vie et en particulier dans l'enfance, et cela en dépit des progrès permis par l'amélioration des conditions d'hygiène, la généralisation de la vaccination et l'introduction de la chimiothérapie microbienne, et quelle que soit l'aire géographique concernée (pays riches ou pays pauvres). Cette distribution épidémiologique récurrente de la mortalité des maladies infectieuses semble donc suggérer que ce sont des mécanismes génétiques différents qui entrent en ligne de compte à différents âges de la vie. Un faisceau de preuves, directes ou indirectes, conforte cette hypothèse d'une architecture génétique de la maladie dépendante de l'âge (Alcaïs, Quintana-Murci, et al., 2010).

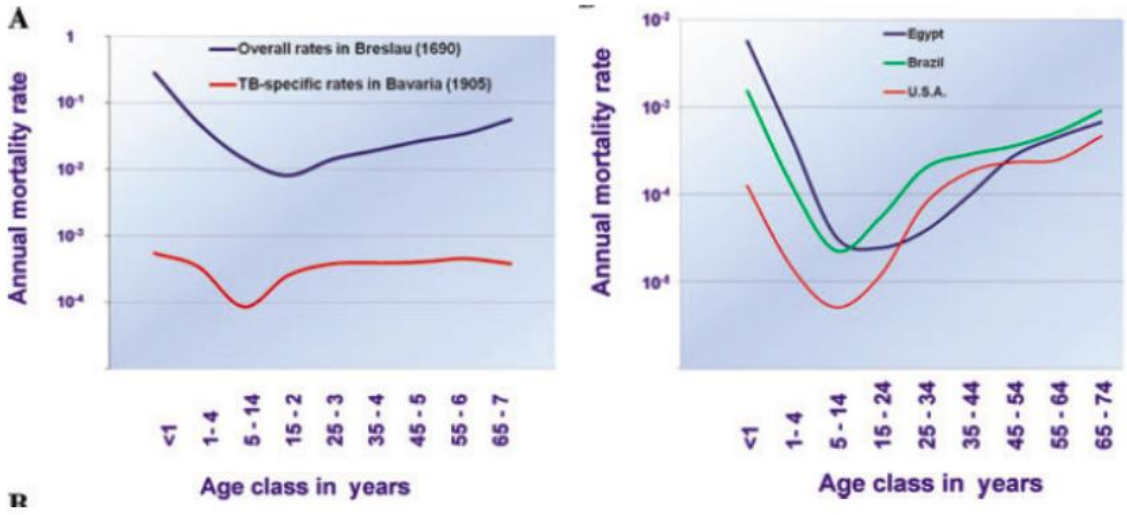


Figure 26 : Taux de mortalité annuelle (en échelle algorithmique) comme une fonction de l'âge avant (A) et après (B) le développement de l'hygiène, la vaccination et la chimiothérapie microbienne (Alcaïs, Quintana-Murci, et al., 2010, p. 26). Sur la figure 26A, la ligne bleue correspond au taux de mortalité globale observée à Breslau par l'astronome anglais Edmund Halley en 1690. La ligne rouge représente le taux de mortalité due à la tuberculose pour les habitants de la Bavière, Allemagne, en 1905. Sur la figure 26B, sont reportés les taux de la mortalité liée aux maladies infectieuses et parasitiques en 2000, aux États-Unis (ligne verte), au Brésil (ligne rouge) et en Egypte (ligne bleue), tels qu'ils sont rapportés par l'Organisation Mondiale de la Santé. On remarque que les taux de mortalité suivent tous la même courbe en L (en particulier avant le développement des mesures anti-infectieuses et à une époque où l'espérance de vie était moindre) et en U.

Une première preuve indirecte est le résultat d'une simulation de l'impact de deux modes d'hérédité d'une maladie infectieuse fictive sur le nombre de cas en fonction de l'âge. Abel et Casanova ont ainsi choisi d'imaginer une maladie infectieuse fictive avec une prévalence de 1% dans une population d'un million d'individus, en considérant que la moitié des nouveau-nés manifestaient une prédisposition mendélienne ou à gène majeur (avec une pénétrance au niveau individuel soit de 1, soit de 0.5) et que la seconde moitié des nouveau-nés étaient atteints de prédisposition polygénique. En admettant que la probabilité d'être infecté est de 0.5 pour chaque individu, on aboutit à trois courbes définissant le nombre de cas en fonction de l'âge : ce sont les courbes représentant le nombre de cas d'individus infectés manifestant une prédisposition mendélienne (avec une pénétrance respective de 1 ou de 0.5) qui correspondent le mieux aux courbes de mortalité précédemment évoquées.

À l'opposé de cette confirmation indirecte et théorique, la génétique a permis de mettre en évidence plusieurs exemples de prédispositions mendéliennes ou d'effet gène majeur / résistance aux infections sévères chez les enfants. C'est le cas de la

susceptibilité à l'encéphalite herpétique que nous avons déjà évoquée, de la septicémie à pneumocoque chez des enfants avec une déficience IRAK-4 et MyD88, qui sont plutôt des maladies rares. Mais c'est aussi le cas d'infections communes comme la tuberculose par exemple que nous avons déjà évoquée et dans laquelle plusieurs mécanismes de susceptibilité mendélienne ou de susceptibilité à gène majeur à la tuberculose ont été mis en évidence chez les enfants ou les très jeunes adultes. En revanche, on sait peu de choses sur les mécanismes génétiques à l'œuvre dans la susceptibilité aux maladies infectieuses chez les adultes, à l'exception de la tuberculose, maladie qui chez l'adulte obéirait probablement à un mécanisme complexe de prédisposition polygénique. La tuberculose ne présente d'ailleurs pas les mêmes caractéristiques cliniques chez l'adulte (tuberculose pulmonaire de réactivation ou infection secondaire d'évolution favorable sous traitement) que chez l'enfant (fréquence des tuberculoses disséminées mortelles), ce que montre le graphique ci-dessous (Figure 27) où le taux de mortalité de la tuberculose est représenté en fonction de l'âge et où les formes cliniques de tuberculose sont représentées en bleu pour la tuberculose disséminée et en rose pour la tuberculose pulmonaire :

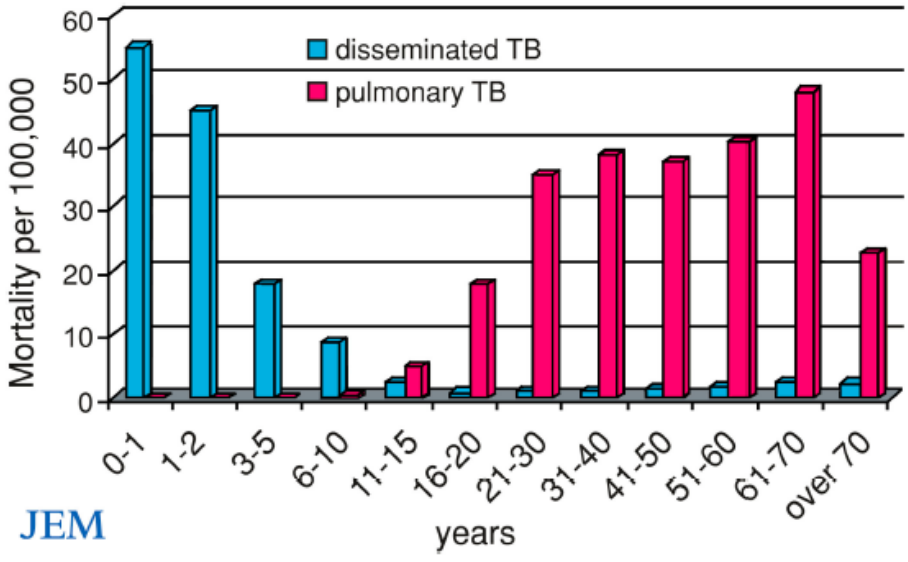


Figure 27 : Taux de mortalité pour la tuberculose disséminée (bleu) et la tuberculose chronique pulmonaire (rose) pour 100 000 individus non traités d'âge varié en Bavière en 1905 (Alcaïs, Fieschi, *et al.*, 2005, p. 1618)

Cette hypothèse sur la théorie génétique des maladies infectieuses est d'ailleurs en cohérence avec l'ensemble des connaissances en immunologie et en particulier avec les différences entre primo-infections et infections secondaires ou réactivations et les différences entre la réponse immunitaire innée et la réponse immunitaire adaptative. En effet, les maladies infectieuses de l'enfance sont le plus généralement des primo-infections, alors que les maladies infectieuses à l'âge adulte correspondent plutôt soit à des infections secondaires, soit à la réactivation d'une infection latente. Or, la réponse immunitaire se décompose en général en trois temps : la réponse immunitaire immédiate, non spécifique, qui fait appel aux cellules non-hématopoïétiques du système immunitaire et qui joue un rôle décisif dans le recrutement de la réponse immunitaire intermédiaire (cellules myéloïdes) et dans le déclenchement de la réponse adaptative (cellules lymphoïdes) qui sont spécifiques du pathogène. Les réponses immunitaires immédiate et intermédiaire jouent donc un rôle primordial dans la réponse aux primo-infections, tandis que la réponse adaptative qui ne se met en place qu'au bout de quelques jours et qui est dépendante de la réponse immunitaire immédiate, est déterminante dans la réponse aux infections secondaires et dans le contrôle des réactivations des infections latentes. Par ailleurs, la réponse immunitaire innée met en jeu un plus petit nombre de gènes que la réponse immunitaire adaptative qui nécessite une recombinaison somatique des gènes. La réponse immunitaire adaptative acquiert également une redondance fonctionnelle au fil des infections, alors que la réponse immunitaire innée n'évolue pas. Ainsi, si les enfants présentant des prédispositions mendéliennes à de multiples infections dues à des défauts majeurs des cellules hématopoïétiques de l'immunité adaptative (les immunodéficiences primaires conventionnelles) aggravent leur état clinique dans le temps, les enfants présentant des prédispositions mendéliennes à une seule infection ou à une seule famille d'infections dues à des défauts majeurs des cellules non-hématopoïétiques de l'immunité innée (les nouvelles immunodéficiences primaires), ont tendance à s'améliorer dans le temps. C'est ainsi que les enfants affectés par une déficience de l'axe TLR3-UNC-93B (prédisposition mendélienne à l'encéphalite herpétique) ne font pas d'épisodes de réactivation, ou en font de moins en moins avec le temps (Casanova et Abel, 2007b). En effet, ces défauts mendéliens de l'immunité innée peuvent être progressivement compensés par la mise en place d'une immunité

acquise et spécifique du pathogène incriminé, alors que les défauts mendéliens de l'immunité acquise ne sont jamais compensés et que les défauts non mendéliens de l'immunité acquise ont un impact moindre sur la sévérité de la primo-infection mais ont un impact décisif sur le contrôle des infections secondaires et des réactivations. Nous avons récapitulé ces différents éléments dans le Tableau 8.

	Prédisposition mendélienne à plusieurs infections (immunodéficiences primaires classiques)	**Prédisposition mendélienne à une infection (nouvelles immunodéficiences primaires)**	**Prédisposition polygénique à une ou plusieurs infections**
Type de réponse immunitaire affectée	Réponse immunitaire adaptative	Réponse immunitaire innée	Réponse immunitaire adaptative
Type de cellules du système immunitaire affectées	Cellules hématopoïétiques lymphoïdes	Cellules non hématopoïétiques	Cellules hématopoïétiques lymphoïdes
Type d'infection	Primo-infection sévère, réactivations et infections secondaires sévères et fréquentes	Primo-infection sévère, réactivations et infections secondaires rares et moins sévères	Primo-infection asymptomatique, réactivations et infections secondaires plus sévères
Apparition de la maladie	Apparition dans l'enfance	Apparition dans l'enfance	Apparition à l'âge adulte
Évolution clinique dans le temps	Dégradation: multiplication des infections opportunistes	Amélioration de l'état clinique: compensation des défauts du système immunitaire inné par le développement de la réponse immunitaire acquise spécifique	Évolution clinique variable

Tableau 8 : Cohérence entre l'hypothèse de l'architecture génétique âge-dépendante des maladies infectieuses et les connaissances en immunologie

L'hypothèse de l'architecture génétique âge-dépendante des maladies infectieuses est également cohérente avec l'ensemble des connaissances de génétique clinique sur les autres classes de maladie : la majorité des défauts monogéniques se manifestent dans l'enfance, tandis que la majorité des maladies correspondantes chez

les adultes se manifestent plus tard. C'est le cas du diabète, de l'obésité et de bien d'autres maladies communes :

> « Cette organisation générale s'applique potentiellement à tous les champs de la médecine […] . Les cinq désordres métaboliques qui peuvent frapper à tout âge – le diabète, l'obésité, l'hypertension, l'ostéoporose et l'athérosclérose – manifestent tous un double mode d'hérédité. Ces maladies sont en général causées par des défauts monogéniques dans la prime enfance, tandis que le mode d'hérédité semble devenir plus complexe au fur et à mesure que l'âge augmente. Par exemple, le diabète néonatal et même le diabète de type 1 avant l'âge de six mois, peuvent être monogéniques, alors que chez des enfants plus âgés, le diabète est généralement dû à un effet de gène majeur. De la même manière, le diabète de type 2 chez les jeunes adultes est souvent monogénique, alors que le diabète de type 2 à déclenchement tardif est apparemment plus complexe. De la même manière, des causes monogéniques de l'obésité, de l'ostéoporose, de l'hypertension, de l'hypercholestérolémie et de l'athérosclérose ont été identifiées chez des enfants tandis que les syndromes correspondants chez les adultes démontrent un mode d'hérédité plus complexe ». (Alcaïs, Quintana-Murci, *et al.,* 2010, p. 27)

Enfin, dernier argument en faveur de cette hypothèse : l'architecture génétique âge-dépendante des maladies infectieuses est cohérente avec les données empiriques sur la différence de pressions sélectives qui s'exercent sur les défauts monogéniques et sur les défauts polygéniques. En effet, les variants alléliques polygéniques ont généralement une pénétrance faible et un impact faible en termes de fitness : ce sont donc des variants alléliques transmis par des individus en âge de reproduction ce qui leur permet d'atteindre une fréquence plus haute dans la population générale que les défauts monogéniques, dont la prévalence et l'impact en termes de fitness sont plus élevés et qui par conséquent sont soumis à une sélection purifiante.

La théorie génétique des maladies infectieuses s'appuie donc sur l'identification de quatre mécanismes génétiques qui ne sont pas spécifiques d'une classe de maladies infectieuses, mais qui sont spécifiques de sous-groupes de populations atteintes d'une maladie infectieuse donnée. Ainsi dans l'ensemble des individus atteints de la tuberculose, on pourra distinguer différentes formes cliniques

correspondant à des populations d'âges différents et à des mécanismes génétiques différents. La théorie génétique des maladies infectieuses aboutit donc d'une part à une nouvelle conception du continuum génétique des maladies infectieuses et d'autre part à une hypothèse empiriquement testable qui prédit que l'architecture génétique d'une maladie infectieuse est dépendante de l'âge. À présent que nous avons présenté les caractéristiques de la théorie génétique des maladies infectieuses, il nous faut analyser en quel sens c'est un exemple de théorie génétique régionale, en nous appuyant sur la liste de conditions que nous avons établie dans la deuxième partie de ce travail.

5.3. En quoi la théorie génétique des maladies infectieuses est-elle un exemple de théorie génétique régionale ?

Pour pouvoir parler d'une théorie génétique régionale, nous avons identifié dans la deuxième partie de ce travail plusieurs conditions : (1) Une théorie génétique doit être intégrée à une explication interactionniste co-constructionniste des maladies et se présente comme une théorie partielle dont l'objectif est d'expliquer le rôle des gènes dans une classe de maladie. (2) a) Une théorie génétique régionale concerne au minimum une des quatre cibles explicatives d'une classe de maladie donnée (symptômes, évolution), b) met en jeu au moins un type d'explication (physiopathologique, épidémiologique ou évolutionnaire) et c) gagne en crédibilité si elle parvient à expliquer la variabilité interindividuelle, la variabilité intra-individuelle ou la comorbidité. Nous allons maintenant passer en revue ces critères afin de déterminer la puissance de la théorie génétique des maladies infectieuses.

5.3.1. La TGMI est-elle une théorie partielle qui s'intègre dans une explication interactionniste co-constructionniste des maladies infectieuses ?

La théorie génétique des maladies infectieuses est présentée par Jean-Laurent Casanova et Laurent Abel comme une théorie partielle des maladies infectieuses : il ne s'agit pas de prétendre expliquer de façon exhaustive la classe des maladies infectieuses, mais seulement d'expliciter le rôle des gènes (de l'hôte) dans l'explication

des maladies infectieuses. Comme nous l'avons rappelé au début du chapitre, Casanova et Abel considèrent d'ailleurs que la théorie génétique des maladies infectieuses est complémentaire de la théorie immunologique, écologique et microbiologique des maladies infectieuses qui éclairent d'autres aspects de l'explication des maladies infectieuses (respectivement le rôle du système immunitaire de l'hôte, de l'ensemble des facteurs environnementaux non-microbiens et des facteurs génétiques et non-génétiques dépendants du micro-organisme). Ils insistent également sur le fait que ces quatre théories partielles font sans cesse référence l'une à l'autre et qu'elles ne se comprennent correctement qu'en référence aux autres théories partielles de l'explication des maladies infectieuses (voir Figure 28) :

> « Nous voudrions aussi insister sur le fait que les théories microbiologique, écologique, immunologique et génétique qui peuvent sembler en compétition, ne sont en réalité pas seulement complémentaires, mais aussi chevauchantes. Par exemple, la théorie immunologique fait référence à des processus somatiques qui restent sous le contrôle de la lignée germinale ; la théorie génétique est une théorie de l'immunité de la lignée germinale et la théorie écologique implique souvent des réponses immunitaires innées ou adaptatives. » (Casanova et Abel, 2013, p. 215)

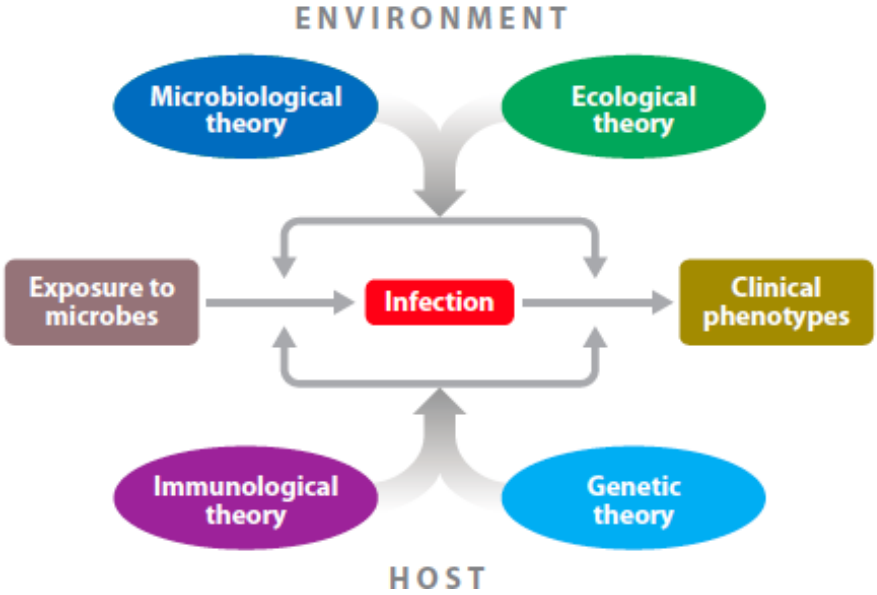


Figure 28 : Les quatre théories complémentaires des maladies infectieuses. (Casanova et Abel, 2013, p. 216). Nous traduisons ici la légende originale de cette figure : « En principe, la variabilité interindividuelle de la présentation clinique (qui va des infections asymptomatiques aux infections fatales) chez les individus infectés dépend de quatre forces intriquées, correspondant aux théories microbiologique, écologique, immunologique et génétique. La maladie est attribuée aux variations microbiennes (qualitative ou quantitative) dans la théorie microbiologique, aux variations environnementales (autres que celles qui concernent le pathogène incriminé) dans la théorie écologique, à un déficit de l'immunité acquise somatique, l'immunité adaptative (à la fois des éléments génétiques et épigénétiques) dans la théorie immunologique et aux erreurs innées de l'immunité encodée par la lignée germinale (qu'elle soit intrinsèque, innée ou adaptative) dans la théorie génétique. Ces quatre théories sont à la fois complémentaires et chevauchantes. »

La théorie génétique des maladies infectieuses est donc bien une théorie partielle qui se focalise sur le rôle d'un facteur causal dans l'explication des maladies infectieuses, en l'occurrence le rôle des gènes de l'hôte ; mais prend-elle place dans un cadre interactionniste ? Indiscutablement, et pour plusieurs raisons. D'abord, comme nous l'avons évoqué dès l'introduction de ce chapitre, les maladies infectieuses sont des exemples paradigmatiques de maladies environnementales, puisque par définition elles impliquent toujours l'action d'un agent infectieux extérieur. Mais surtout, la matrice causale que décrivent Casanova et Abel et qu'on retrouve dans la Figure 29, est une matrice causale complexe qui ne se contente pas de définir une opposition binaire entre gènes et environnement mais qui décrit au moins quatre catégories de causes qui interagissent ensemble pour expliquer l'apparition et le développement des maladies infectieuses : du côté de l'hôte, les facteurs génétiques et les facteurs non génétiques, du côté de l'environnement les facteurs microbiens (parmi lesquels on

retrouve les facteurs génétiques et les facteurs non-génétiques de l'agent infectieux) et les facteurs non-microbiens (le climat, les conditions socio-économiques, etc.)

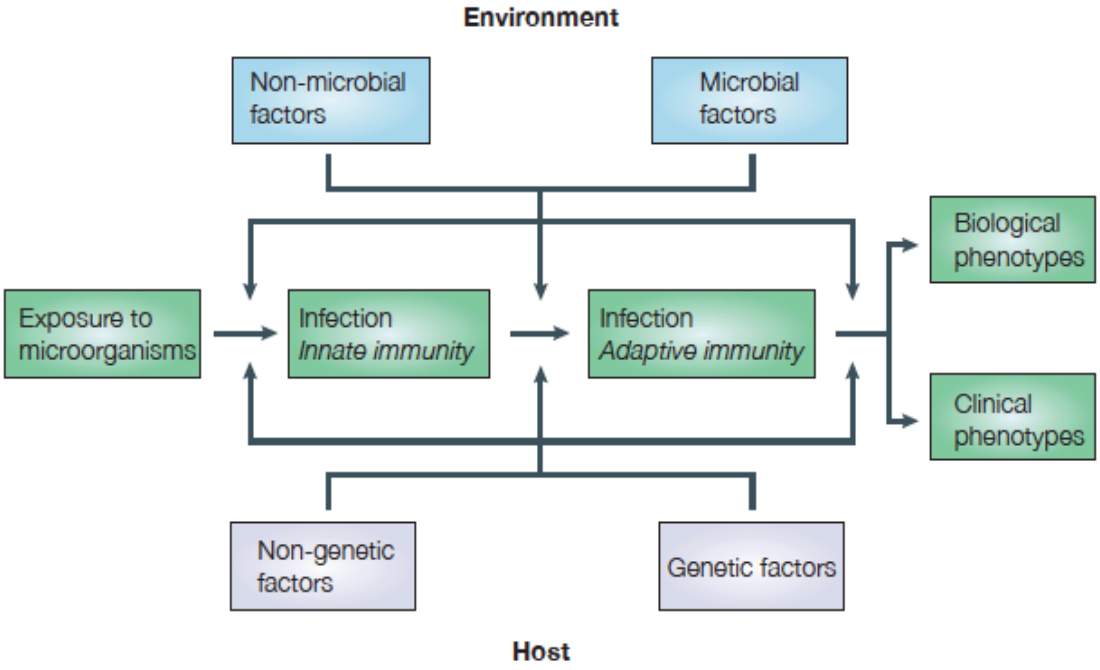


Figure 29 : Représentation schématique des interactions hôte-environnement au cours d'une infection (Casanova et Abel, 2004, 56). L'exposition à un agent infectieux n'aboutit pas toujours au développement clinique d'une maladie infectieuse. Le développement de l'infection dépend de l'immunité innée et/ou de l'immunité adaptative. L'infection peut être détectable biologiquement ou cliniquement. Le complexe processus qui va de l'exposition à l'agent infectieux au développement clinique des maladies infectieuses, est sous le contrôle de facteurs dépendant de l'hôte, génétiques (par exemple, la mutation dans un gène impliqué dans l'immunité à l'infection) ou non-génétiques (une plaie cutanée, par exemple), qui peuvent influencer l'exposition au micro-organisme (barrières cutanéo-muqueuses) ou l'infection par le micro-organisme (immunités innée et adaptative) et de facteurs environnementaux microbiens (comme les facteurs de virulence), ou liés au mode d'exposition (par exemple, la température de l'air) qui peuvent avoir un impact à chaque étape de l'interaction.

La théorie génétique des maladies infectieuses est donc une théorie partielle du rôle des gènes de l'hôte dans l'explication des maladies infectieuses et prend place dans un cadre interactionniste. Mais de quel cadre interactionniste s'agit-il ? La réponse à cette question ne peut être qu'ambiguë. En un sens, la théorie génétique des maladies infectieuses s'appuie sur de nombreux concepts issus de l'interactionnisme quantitatif, dont nous avons discuté l'importance en génétique humaine. Des concepts comme ceux de « gène majeur » ou de « prédisposition polygénique » sont typiquement des concepts issus de la génétique des populations et nés dans le contexte de l'analyse de ségrégation complexe, c'est-à-dire en se fondant

sur des techniques statistiques de dissection de la génétique des maladies mendéliennes et non mendéliennes, qui sont intrinsèquement liées aux notions d'héritabilité et d'architecture génétique de la maladie.

Cependant, ce sont bien des interactions qualitatives qui sont présentées dans la Figure 29 : interactions entre la génétique de l'hôte et celle du parasite, etc. Il semble néanmoins que certaines interactions ne soient pas représentées ou que certaines interactions soient grossièrement simplifiées : par exemple, l'influence de l'hôte sur son environnement et rétroactivement l'influence de l'environnement modifiée par l'hôte sur la génétique de la maladie ne sont pas représentées. Cette ambiguïté se retrouve dans la représentation que proposent Casanova et Abel des rapports entre susceptibilité / vulnérabilité génétique de l'hôte aux infections et virulence du microbe (Figure 30). Comme nous l'avons mentionné un peu plus haut, le concept de virulence a longtemps été considéré soit comme un concept intrinsèquement lié au micro-organisme (qui serait doté de propriétés intrinsèques comme la notion d'îlots de pathogénicité et de facteurs de virulence qui sont responsables de son caractère pathogène) soit comme un concept intrinsèquement lié à l'hôte (et à sa capacité de réponse immunitaire). Or aucune de ces conceptions de la virulence ne suffit à rendre compte de la multiplicité des interactions possibles entre hôte et micro-organisme qui comprennent des situations aussi différentes que le mutualisme, le commensalisme, l'opportunisme, le parasitisme, les infections asymptomatiques, les infections latentes, les infections létales, etc. Ainsi, Casadevall et Pirofsky qui sont les hérauts d'une compréhension plus interactionniste du concept de virulence, proposent un cadre interactionniste et co-constructionniste pour comprendre la nature des interactions entre l'hôte et le microbe qu'ils appellent « le cadre de la réponse aux dommages » de la pathogénèse microbienne (*Damage-response Framework*) :

> « Le cadre de la « réponse aux dommages » de la pathogénèse microbienne est fondé sur trois piliers. Tout d'abord, la pathogénèse microbienne est le résultat d'une interaction entre un hôte et un micro-organisme et n'est pas seulement attribuable ni au micro-organisme ni à l'hôte. Deuxièmement, le résultat pathologique de l'interaction hôte-micro-organisme est déterminé par le niveau de dommage infligé à l'hôte. Troisièmement, le dommage infligé à

> l'hôte résulte de facteurs microbiens et/ou de la réponse de l'hôte. » (Casadevall et Pirofski, 2003, p. 17)

Le cadre proposé par Casadevall et Pirofski est plus complexe que ces trois affirmations : ces deux auteurs ont ainsi développé une classification des micro-organismes en six classes (Casadevall et Pirofski, 1999, 2000) et ont redéfini les concepts de pathogène et de virulence, en essayent de prendre en compte les deux dimensions (hôte-centré et micro-organisme centré) :

> « Le cadre de la réponse aux dommages définit la virulence comme la capacité relative d'un micro-organisme à causer des dommages à un hôte. Le terme « relatif » est inclus parce que la virulence est fréquemment mesurée en comparaison d'un autre micro-organisme ou d'un variant du même micro-organisme. Cependant, lorsqu'une compréhension plus complète des mesures quantitatives des dommages est disponible, il devrait être possible de se dispenser du terme « relatif ». Les définitions du pathogène et de la virulence dans le cadre de la réponse aux dommages montrent bien que c'est seulement dans un hôte susceptible qu'un micro-organisme est pathogène et que la virulence peut s'exprimer. En conséquence, ni les caractéristiques d'un pathogène, ni la virulence ne peuvent être considérées comme des variables microbiennes indépendantes. (Casadevall et Pirofski, 2003, p. 19)»

Cette redéfinition des concepts de virulence et de pathogène en termes d'interactions semble compatible avec la façon dont Casanova et Abel conçoivent le développement des maladies infectieuses (voir Figure 30), qui résulte de l'interaction entre la génétique de l'hôte et la génétique du parasite et qu'ils commentent ainsi : *« Un individu avec une vulnérabilité génétique forte (c'est-à-dire due à un variant monogénique) pourra développer une maladie à la suite d'une infection par un microbe faiblement virulent, tandis qu'un individu avec un faible niveau de vulnérabilité génétique pourra développer une maladie infectieuse si, et seulement si, il est infecté par un microbe hautement virulent. »* (Casanova et Abel, 2013, p. 223). Cependant, et c'est là où résulte l'ambiguïté de l'interactionnisme que défendent Casanova et Abel, la virulence apparaît dans leur schéma comme une propriété du microbe et l'ensemble

des autres interactions (rôle de l'environnement non-microbien, etc.) n'apparaissent pas comme des déterminants de l'infection.

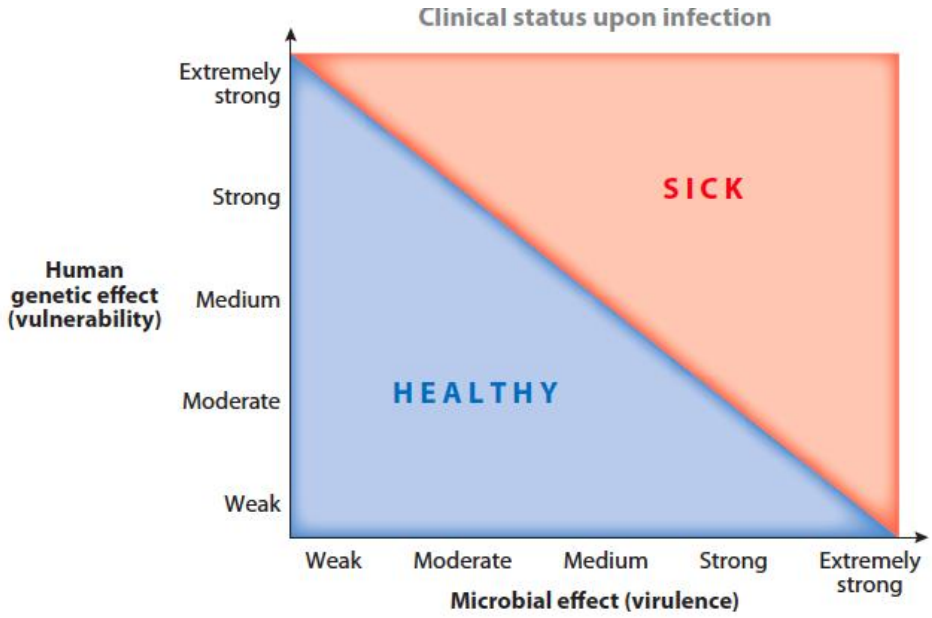


Figure 30 : Contributions respectives de la génétique de l'hôte et de la génétique du microbe dans le développement clinique des maladies infectieuses (Casanova et Abel, 2013, p. 223).

Pour résumer, la théorie génétique des maladies infectieuses peut être considérée comme une théorie partielle centrée sur l'explication du rôle de la génétique de l'hôte dans les maladies infectieuses mais qui s'intègre en principe à trois autres théories contribuant à l'explication des maladies infectieuses (la théorie microbiologique, la théorie écologique, la théorie immunologique) et dans une matrice causale complexe qui suppose des interactions entre la génétique de l'hôte, les facteurs non génétiques de l'hôte, le microbe et l'environnement. Néanmoins, la théorie génétique des maladies infectieuses s'appuie également sur des concepts de génétique des populations qui sont intrinsèquement liés à un interactionnisme de type quantitatif et l'intégration de la théorie génétique des maladies infectieuses aux autres explications des maladies infectieuses reste souvent évasive.

Enfin, la théorie génétique des maladies infectieuses ne fait jamais allusion à l'existence du microbiome, c'est-à-dire à l'existence d'une flore microbienne commensale à l'homme qui influence notre réponse immunitaire aux infections et qui est potentiellement un autre facteur génétique déterminant du développement des

maladies infectieuses. En effet, ce qu'implique le microbiome c'est que l'être humain n'est pas un organisme « monogénomique » mais au contraire « multigénomique », et que ce n'est pas indépendamment de la génétique du microbiome que l'hôte réagit à son environnement (Dupré, 2011). Autrement dit, si la théorie génétique des maladies infectieuses s'intègre au moins en principe dans un cadre interactionniste co-constructionniste, ce cadre est encore peu développé et il serait intéressant de discuter à quel point cette théorie génétique des maladies infectieuses est compatible avec les avancées récentes et à venir issues de notre connaissance du rôle du microbiome dans le développement des maladies infectieuses.

5.3.2. La TGMI est-elle une théorie régionale qui explique les quatre cibles explicatives des maladies infectieuses ?

Pour parler d'une théorie génétique d'une classe de maladie, nous avons établi que la théorie doit expliquer le rôle des gènes dans au moins l'une des quatre cibles explicatives d'une classe de maladie : les causes, les symptômes, l'évolution ou le traitement. La théorie génétique des maladies infectieuses intervient sur ces quatre cibles explicatives, aussi bien du point de vue de la classe de maladie (les maladies infectieuses) que pour une maladie type (ex : la lèpre, la tuberculose, etc.). Ainsi, la théorie génétique des maladies infectieuses explique :

(1) Une cause des maladies infectieuses. En tant que théorie partielle, la théorie génétique des maladies infectieuses explique par quels mécanismes la génétique de l'hôte cause les maladies infectieuses. Les quatre mécanismes génétiques précédemment présentés sont autant de façons dont les gènes de l'hôte causent la présence ou l'absence de maladies infectieuses. Chaque mécanisme génétique est considéré comme le mécanisme qui fait la différence dans une population donnée lorsque toutes les autres conditions (les conditions environnementales, les conditions d'exposition à une même souche microbienne, absence de vaccination, etc.) sont fixées. Dans ces conditions, c'est le fait de porter tel variant allélique prédisposant de façon mendélienne à tel micro-organisme pathogène qui explique l'apparition de la maladie chez un individu X tandis que l'individu Y, qui a été exposé dans les mêmes conditions à

la même souche microbienne et qui n'est pas non plus vacciné, ne développe pas la maladie et est soit résistant (s'il est porteur d'un variant allélique qui entraine une résistance à l'infection), soit porteur asymptomatique, etc.

(2) Les symptômes des maladies infectieuses. Différentes catégories de symptômes peuvent être expliquées par la théorie génétique des maladies infectieuses : variabilité des formes cliniques (différents types de symptômes pour une même maladie), variabilité de la sévérité (formes asymptomatiques, formes « classiques » modérément sévères, formes très sévères potentiellement fatales) et de l'âge de déclenchement des symptômes (primo-infections potentiellement fatales chez les enfants et réactivations / infections secondaires plus modérées chez les adultes). Ainsi, dans le cas de la lèpre, deux groupes de variants alléliques ont été mis en évidence (Alcaïs, Mira, *et al.,* 2005 ; Alter *et al.,* 2011 ; Gaschignard *et al.,* 2013), réalisant un modèle de la susceptibilité génétique à la lèpre en deux étapes (voir Figure 31) : le premier groupe de variants alléliques est impliqué dans le déclenchement de la lèpre *per se*, c'est-à-dire dans le fait de développer des symptômes de la lèpre après avoir été exposé au bacille de Hansen et le second groupe de variants alléliques conditionne le type de réponse immunitaire que l'hôte développe vis-à-vis de *Mycobacterium leprae*, conduisant à des symptômes différents en fonction des variants alléliques (forme à lésion unique, forme paucibacillaire, forme multibacillaire).

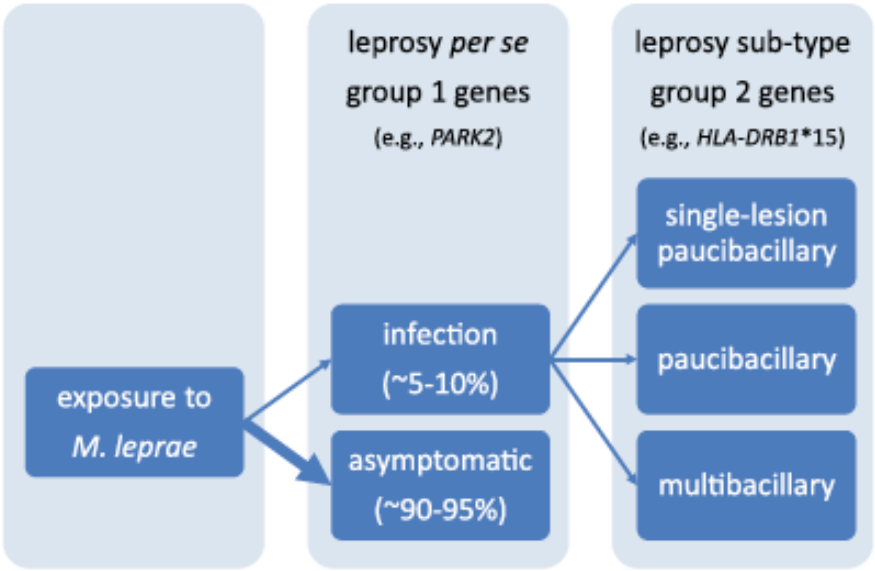


Figure 31 : Modèle en deux étapes de la susceptibilité génétique à la lèpre (Alter et al., 2011, p. 25). Suite à l'exposition à *Mycobacterium leprae*, le groupe 1 de gènes confère une susceptibilité (ou une résistance) à l'établissement de l'infection et au développement clinique de la maladie (la lèpre *per se*). Parmi les 5 à 10% d'individus qui vont développer la lèpre, le groupe 2 de gènes détermine le type de réponse immunitaire de l'hôte et par conséquent le type de lésions cliniques qui vont se développer.

L'exemple de la tuberculose est également éclairant : comme nous l'avons dit précédemment, l'exposition à *Mycobacterium Tuberculosis* aboutit rarement au développement clinique de la tuberculose, puisqu'on estime que 90% des individus infectés restent asymptomatiques toute leur vie. Parmi les 10 % de sujets infectés et malades, il y a une nette distinction entre les formes de l'enfance, qui sont des formes extrêmement sévères et disséminées de la tuberculose et les formes de l'adulte, qui sont des formes moins sévères et plus localisées (tuberculose pulmonaire chronique dans la majorité des cas). C'est le type de mécanisme génétique impliqué qui explique cette distribution âge-dépendante des formes cliniques et de la sévérité de la maladie en deux tranches d'âge, à tel point que Casanova et Abel parlent de la tuberculose infantile et de la tuberculose de l'adulte comme « deux maladies distinctes » (Alcaïs, Fieschi, *et al.,* 2005).

(3) L'évolution de la maladie. La TGMI explique l'évolution des maladies infectieuses en deux sens. Elle explique d'une part l'évolution des symptômes

de la maladie sans traitement (évolution sévère et souvent fatale chez les formes infantiles non traitées de prédisposition mendélienne à une infection et à de multiples infections, évolution plus favorable chez les formes du jeune adulte avec gène majeur de susceptibilité à une infection ou chez les adultes manifestant une prédisposition polygénique à une infection) et d'autre part l'évolution de la maladie dans le temps (risque de récurrence très élevé chez les immunodéficiences primaires conventionnelles, alors que les patients atteints de « nouvelles » immunodéficiences primaires font très peu de récurrence et que les formes récurrentes sont moins graves que la primo-infection étant donné la mise en place d'une immunité secondaire acquise).

(4) Le traitement des maladies infectieuses. En fonction du mécanisme génétique, ce ne sont pas les mêmes traitements qui sont suggérés. Ainsi, pour les prédispositions mendéliennes à de multiples infections qui atteignent les cellules hématopoïétiques de façon ciblée, la thérapie génique consistant à transférer un gène fonctionnel dans un vecteur dans les cellules du patient a été proposée dès les années 1990 par l'équipe d'Alain Fischer dans le cadre du syndrome d'immunodéficience sévère combinée (Hacein-Bey-Abina *et al.,* 2010). La greffe de moelle est une autre option thérapeutique largement utilisée dans ce cadre. Pour les nouvelles immunodéficiences primaires en revanche, des médicaments immunostimulateurs correspondant à la voie de l'immunité naturelle touchée, ont été développés : facteurs de croissance stimulant la croissance des granulocytes (*Granulocyte-Colony Stimulating Factor*) dans la neutropénie congénitale sévère, interféron gamma dans les prédispositions mendéliennes aux mycobactéries, interféron alpha pour les prédispositions mendéliennes aux encéphalites herpétiques (Casanova et Abel, 2007b ; Casanova et al., 2005).

5.3.3. Quel type d'explication mobilise la TGMI ?

5.3.3.1. Rappels sur l'explication mécaniste

Quel type d'explication mobilise la théorie génétique des maladies infectieuses ? Casanova et Abel font explicitement allusion à quatre « mécanismes » génétiques, mais le choix de ce terme peut interroger quand on connaît les débats actuels sur la définition des mécanismes biologiques et leur pertinence dans les explications biologiques (Bechtel et Abrahamsen, 2005 ; Campaner, 2011). Si on considère en général que les débats contemporains sur la place des mécanismes dans les explications biologiques ont commencé dès la publication de l'ouvrage de Bechtel et Richardson (Bechtel et Richardson, 1993), c'est la caractérisation des mécanismes proposée par Machamer, Darden et Craver dans leur célèbre article « Thinking about mechanisms » qui est devenue une référence dans la communauté philosophique ces dernières années, supplantant les autres tentatives de définitions des mécanismes (Machamer et al., 2000). MDC caractérisent les mécanismes biologiques comme « des entités et des activités organisés de façon à produire des changements réguliers à partir de conditions initiales ou de départ jusqu'à des conditions finales » (Machamer ,Darden et Craver*, 2000, p. 3)*. Même si la définition des explications mécanistes ou le statut ontologique de ces mécanismes peuvent différer selon les auteurs (Glennan, 2005 ; Machamer, 2004), on retrouve dans cette définition les caractéristiques essentielles du concept : les explications mécanistes sont des explications causales fondées sur un dualisme ontologique qui implique des parties ou des entités d'une part et des activités ou des opérations d'autre part. C'est la façon dont ces entités et ces activités sont organisées dans un processus temporel continu qui garantit un certain niveau de généralisation de l'explication mécaniste et qui permet de fournir des explications dynamiques des phénomènes, un problème fréquent en biologie. En somme, pour reprendre les termes de Dominic Murphy : *« Bien que les philosophes ne soient pas d'accord sur la compréhension précise de la nature des mécanismes, il existe un consensus sur le fait que lorsqu'on rend compte de la façon dont un mécanisme explique un phénomène, il faut mentionner (i) des parties*

physiques qui (ii) font des choses » (Murphy, 2010, p. 602). Étant donné cette caractérisation des mécanismes, l'usage de ce terme dans la théorie génétique des maladies infectieuses peut soulever deux objections : la première sur le niveau de détail et la seconde sur les fondements conceptuels des mécanismes suggérés.

5.3.3.2. Première objection – des mécanismes insuffisamment détaillés ?

En effet, dans une certaine mesure, la théorie génétique des maladies infectieuses mobilise le type d'entités exigé par la caractérisation MDC, sous la forme de variants alléliques ou de maladies. Il semble aussi que pour chaque mécanisme, il y ait bien une organisation régulière entre des variants alléliques et le développement des infections. Cependant, ce ne sont pas les entités et les formes d'organisation qu'on attendrait en parlant de « mécanismes génétiques ». Lorsqu'on parle de « mécanismes génétiques », on s'attend à être confronté à des activités moléculaires, telles que la réplication, la transcription et la traduction de l'ADN, la régulation de l'expression génétique, etc. Autrement dit, ces mécanismes sont-ils suffisamment détaillés pour permettre la description d'un processus temporel continu, consubstantiel à la notion de mécanisme ?

Cette première objection n'est pas tant un obstacle à une description mécaniste de la théorie génétique des maladies infectieuses qu'un problème de degré d'explication – les mécanismes moléculaires ne sont pas tant absents qu'implicites. Dans la mesure où la théorie génétique des maladies infectieuses décrit des mécanismes généraux s'appliquant à une classe de maladie, il y a nécessairement un compromis (*trade-off*) à trouver entre le niveau de généralité du mécanisme et le niveau de détails et de précision que l'on peut atteindre dans la description. Pour le dire encore différemment, ce que les partisans d'une théorie génétique des maladies infectieuses proposent, ce n'est ni une description complète du niveau moléculaire de chaque mécanisme (auquel cas on s'attendrait à un schéma général pour décrire chaque mécanisme), ni une explication d'un cas particulier de susceptibilité génétique pour une maladie infectieuse donnée (auquel cas les activités et les entités seraient plus précises). D'ailleurs, lorsque Casanova et Abel présentent en détail un cas particulier de susceptibilité génétique pour une maladie donnée, comme l'ensemble

des déficits mendéliens responsables de la susceptibilité à l'encéphalite herpétique, on retrouve le vocabulaire des activités (« induire l'activation », « liaison », « phosphorylation ») et des entités (« ARN double brin », « éléments de réponse de l'ADN », « activateur de transcription », « récepteur transmembranaire », « noyau ») de la biologie moléculaire et de la biologie cellulaire, ainsi que le recours à des schémas permettant de visualiser les différentes étapes du mécanisme (Figure 32).

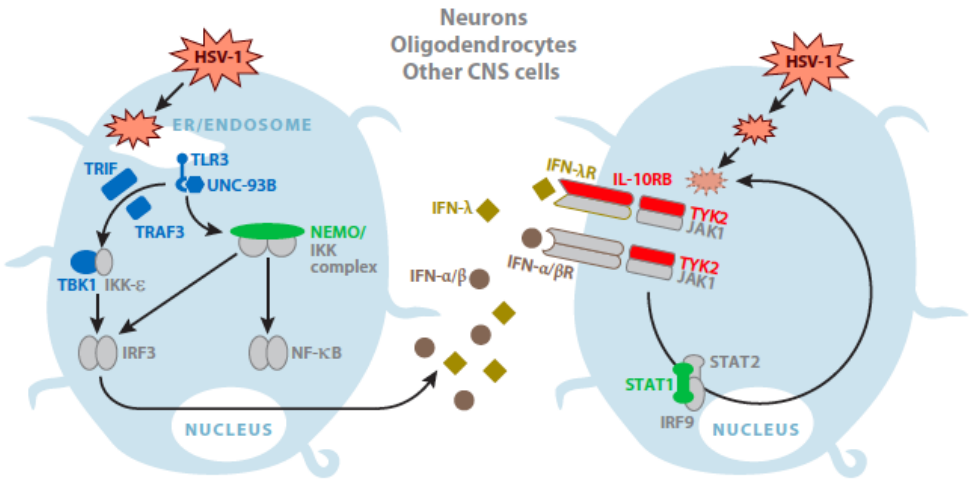


Figure 32 : Erreurs innées de l'immunité dépendante de l'axe TLR3-IFN α/β dans l'encéphalite herpétique infantile (HSE) (Casanova et Abel, 2013, p. 230) : « Le virus HSV1 produit un ARN viral double brin (ARNdb) durant sa réplication. TLR3 est un récepteur transmembranaire de l'ARNdb dans le réticulum endoplasmique et dans l'endosome de la plupart des cellules. Dans les cellules du système central nerveux (SNC), la reconnaissance de l'ARNdb par TLR3 induit l'activation des chemins IRF-3 et NF-κB via TRIF, ce qui entraine la production d'IFN α/β et/ou d'IFN λ. Des mutations sur TLR3, UNC-93B, TRIF, TRAF3, TBK1 et NEMO sont associées à une production défectueuse d'IFN α/β et/ou d'IFN λ, en particulier durant l'infection HSV-1. La liaison d'IFN α/β et/ou d'IFN λ à leurs récepteurs induit la phosphorylation de JAK1 et TYK2, activant la transduction du signal des protéines STAT1, STAT2 et IRF9. Ce complexe est transloqué, sous forme d'un hétérodimère, dans le noyau où il agit comme un activateur de transcription, en se liant aux éléments spécifiques de réponse de l'ADN dans la région du promoteur des gènes induisant l'IFN. Les déficiences en STAT1 sont donc associées à une réponse défectueuse d'IFN α/β et/ou d'IFN λ. Les protéines, pour lesquelles des mutations génétiques ont été identifiées et sont associées à des cas d'encéphalite herpétique isolés, sont représentées en bleu. Les protéines pour lesquelles des mutations ont été identifiées et qui sont associées à des susceptibilités aux infections mycobactériennes, bactériennes et aux maladies virales (et qui incluent l'encéphalite herpétique) sont montrées en vert. Les protéines, pour lesquelles des mutations génétiques ont été identifiées mais qui ne sont pas directement associées à HSE sont représentées en rouge ».

Un autre argument important qui explique le caractère incomplet de la description des mécanismes de la théorie génétique des maladies infectieuses tient au caractère récent de la théorie et à la nature de la recherche scientifique. Dans la mesure où la théorie elle-même est en cours d'élaboration (certains de ces

mécanismes, comme celui de la « résistance mendélienne aux infections », n'ont été décrits que très récemment), la description est nécessairement incomplète. Elle constitue cependant ce que Craver décrirait comme une « esquisse mécaniste ». Une esquisse est une abstraction pour laquelle les entités et les activités du niveau inférieur ne peuvent (pas encore) être décrites ou bien qui contient des « trous » entre ces étapes. Dans la description de la continuité spatiale et temporelle avec laquelle le mécanisme se produit, il y a entre deux étapes des pièces manquantes, des boîtes noires, que nous ne savons pas encore comment compléter (Machamer, Darden et Craver, 2000, p. 18). Ainsi, les mécanismes de la théorie génétique des maladies infectieuses semblent plus proches des esquisses mécanistes que d'une description complète de la susceptibilité génétique aux maladies infectieuses. Cependant, même ces esquisses mécanistes ont un but bien précis : elles constituent des outils heuristiques qui permettent d'indiquer ce qui reste à faire pour obtenir une meilleure explication mécaniste. C'est bien cette dimension heuristique que l'on voit à l'œuvre dans la représentation schématique des défauts mendéliens associés à la susceptibilité génétique à l'encéphalite herpétique (Figure 32) : certaines protéines sont associées à des cas d'encéphalite herpétique (bleu) ou à d'autres maladies infectieuses (verts) tandis que les protéines en rouge désignent des protéines et donc des gènes-candidats qui potentiellement pourraient être associés à une susceptibilité à l'encéphalite herpétique, mais dont l'implication n'a pas encore été prouvée.

5.3.3.3. Deuxième objection – Des mécanismes fondés sur des concepts de l'interactionnisme quantitatif

La deuxième objection est plus sérieuse parce qu'elle concerne les fondements conceptuels des mécanismes décrits par la théorie génétique des maladies infectieuses. Lorsque nous avons discuté le type d'interactionnisme impliqué par la théorie génétique des maladies infectieuses, nous avons souligné l'importance de l'interactionnisme quantitatif dans la formulation de cette théorie. Or, comme les concepts de l'interactionnisme quantitatif ont eux-mêmes été récemment remis en question (comme nous l'avons montré dans le chapitre 3), on peut s'interroger sur la

légitimité d'utiliser ces concepts pour caractériser la théorie génétique des maladies infectieuses.

La description du concept de gène majeur est emblématique de cette difficulté. Le concept de gène majeur est un concept né dans le cadre de l'analyse de ségrégation complexe, c'est-à-dire dans le cadre des méthodes statistiques de la génétique quantitative. C'est un concept à cheval entre le concept de prédisposition mendélienne et la notion de prédisposition polygénique. Or, on a montré dans les chapitres 1 et 2 de cet ouvrage à quel point le concept de maladie mendélienne monogénique était discuté : il est très difficile de trouver un seul exemple de maladie mendélienne au sens strict du terme, puisque même dans les cas paradigmatiques, comme la drépanocytose, la phénylcétonurie ou la mucoviscidose, on a mis en évidence l'existence de phénomènes d'hétérogénéité allélique, d'hétérogénéité génétique, et de gènes modificateurs qui remettent en cause l'équation une mutation – un gène – une maladie. Par ailleurs, on a déjà souligné que l'identification d'une prédisposition vraiment polygénique requiert un large nombre d'individus, à la fois à cause du petit effet attendu attribuable à chaque gène et en raison de la nature additive de ces effets génétiques. Cela peut expliquer pourquoi la preuve que ces mécanismes génétiques ont lieu à la fois au niveau d'une population et au niveau individuel n'a pu être obtenue que par des études de susceptibilité aux maladies infectieuses sur des modèles animaux de maladies infectieuses expérimentalement induites et n'a pas encore été mise en évidence dans les études humaines. Le concept de gène majeur est donc un concept dérivé de l'interactionnisme quantitatif et qui désigne une situation intermédiaire entre deux concepts, qui sont eux-mêmes remis en question par la littérature biomédicale contemporaine.

Par ailleurs, on a souligné dans le chapitre 3 les difficultés rencontrées par le concept de prédisposition : ce concept n'a de sens et n'est opérationnel que si l'on considère qu'il n'y a pas d'interactions gène x environnement, susceptible de venir bouleverser le sens dans lequel se déploie la prédisposition (Tabery, 2009b). Ainsi, dans l'exemple développé par Caspi et Moffitt (Caspi *et al.*, 2002), une interaction gène x environnement vient abolir la notion de prédisposition, puisque les individus qui étaient génétiquement les plus prédisposés à avoir un comportement antisocial dans un environnement familial stable, deviennent les moins prédisposés à un

comportement antisocial dans un environnement de maltraitance familiale (et inversement). Aucune prédisposition de ce type n'a encore été mise en évidence dans le cadre de la théorie génétique des maladies infectieuses *(communication personnelle avec Laurent Abel),* mais comme nous l'avons souligné dans le chapitre 3, de telles interactions sont extrêmement difficiles à détecter par les méthodes statistiques et supposent le recours à d'autres méthodes expérimentales que celle de la génétique quantitative. Cela signifie que le fait de ne pas avoir mis en évidence l'existence d'interactions gène x environnement dans le cadre de la théorie génétique des maladies infectieuses ne permet pas de conclure que de telles interactions n'existent pas.

Cette deuxième objection serait difficile à discuter, si Abel et Casanova n'étaient pas eux-mêmes conscients de cette difficulté. Par exemple, dans un de leurs articles sur la tuberculose, que nous avons mentionné précédemment et qui est intitulé « Tuberculosis in children and adults – two distinct genetic diseases », Abel et Casanova écrivent :

> « […] la tuberculose disséminée semble refléter des prédispositions mendéliennes chez une fraction d'enfants, tandis que la tuberculose pulmonaire adulte semble refléter une prédisposition génétique plus complexe. Cependant, **ces modes d'hérédité sont séparés artificiellement**[50] parce que des gènes modificateurs ont un impact profond sur l'expression de traits mendéliens. Par ailleurs, les gènes majeurs, qui sont des gènes dont les polymorphismes communs exercent un effet significatif suffisamment fort pour être détectés dans des analyses de ségrégation ou dans associations pangénomiques, peuvent exercer un impact presque mendélien. De plus, les prédispositions mendéliennes et les prédispositions complexes ne sont pas exclusives mutuellement dans un groupe d'âge donné, comme cela a été suggéré par la description récente des allèles de susceptibilité NRAMP1 dans les tuberculoses de l'enfant. » (Alcaïs, Fieschi, *et al.*, 2005, p. 1620)

[50] C'est nous qui soulignons.

Le caractère « artificiel » des mécanismes décrits est donc parfaitement assumé par Abel et Casanova : en un sens, il n'y a pas quatre mécanismes de susceptibilité génétique aux maladies infectieuses, mais une infinité, ce qui est d'ailleurs parfaitement cohérent avec la notion de « spectre » ou de « continuum génétique » des maladies infectieuses qu'ils défendent. Ces quatre mécanismes génétiques sont bien des abstractions idéalisées et artificielles de ce qui est réellement à l'œuvre dans la susceptibilité génétique des maladies infectieuses : ils permettent de distinguer quatre situations possibles, dont aucune n'existe véritablement *in natura*, mais qui permettent de comprendre le sens du continuum génétique des maladies infectieuses et d'orienter la recherche sur la susceptibilité génétique des maladies infectieuses.

5.3.3.4. Une explication mécaniste cohérente avec des données évolutionnaires et épidémiologiques

Malgré ses imperfections et ses insuffisances, la théorie génétique des maladies infectieuses fait bien appel à une explication de type mécaniste. Ce n'est cependant pas une explication mécaniste « pure » : comme on l'a vu lorsque nous avons détaillé l'hypothèse selon laquelle l'architecture allélique de la susceptibilité à une maladie infectieuse est dépendante de l'âge, la théorie génétique des maladies infectieuses s'appuie également sur une explication évolutionnaire et sur une explication épidémiologique de la génétique des maladies infectieuses. Ainsi, l'identification des quatre mécanismes génétiques permet d'expliquer à la fois la distribution des courbes de mortalité des maladies infectieuses et la fréquence des allèles morbides dans la population générale et dans des sous-groupes de population. Cette cohérence de l'explication mécaniste physiopathologique avec les explications évolutionnaire et épidémiologique renforce la crédibilité de la théorie génétique des maladies infectieuses.

5.3.4. En quoi la TGMI explique-t-elle la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité des maladies infectieuses ?

Dans le chapitre 4, nous avons identifié trois classes de faits dont l'explication est un gage de précision pour une théorie médicale donnée. Nous avons insisté tout au

long de ce chapitre sur l'objectif initial explicitement réclamé par Abel et Casanova pour la théorie génétique des maladies infectieuses, qui est l'explication de la variabilité interindividuelle aux infections, imparfaitement expliquée jusqu'ici par la théorie des germes, la théorie microbienne, la théorie immunologique et la théorie écologique des maladies infectieuses. Nous nous contenterons donc ici de résumer très rapidement en quoi la théorie génétique des maladies infectieuses remplit cet objectif. Dans une population donnée, pour une maladie donnée, la théorie génétique des maladies infectieuses permet de comprendre 1) le cas des individus exposés non infectés (cas de résistance à une infection), 2) le cas des individus exposés, infectés, mais non symptomatiques (individus immunocompétents pour une infection donnée) 3) le cas des primo-infections sévères à potentiellement mortelles chez l'enfant (prédisposition mendélienne à une infection), chez l'adulte jeune (gène majeur à une infection) et des réactivations / infections secondaires graves de l'adulte (prédisposition polygénique à une infection) 4) le cas de la variabilité clinique des symptômes entre deux individus infectés par la même souche pathogène.

Qu'en est-il de la variabilité intra-individuelle ? Dans le chapitre 4, nous avons défini la variabilité intra-individuelle comme le fait que le même individu ne manifeste pas les mêmes symptômes dans des circonstances similaires. Un exemple de variabilité intra-individuelle est de comprendre comment un individu qui a une prédisposition mendélienne pour une infection donnée n'est pourtant pas susceptible aux réactivations ou infections secondaires, ou en tous cas, aura des symptômes moins sévères en cas de réactivation. La théorie génétique des maladies infectieuses affirme que les prédispositions mendéliennes touchent avant tout les cellules non-hématopoïétiques : une fois le patient infecté et guéri de son infection, la réaction immunitaire adaptative permet la mise en place de défenses immunitaires spécifiques qui permettent de contrôler l'infection.

A quel point la théorie génétique des maladies infectieuses permet-elle d'expliquer la comorbidité des maladies infectieuses ? Une des difficultés que rencontrait la théorie des germes c'est de ne pas pouvoir expliquer pourquoi un individu est sensible à certaines familles d'infections. La théorie génétique des maladies infectieuses représente sans conteste un gain explicatif de ce point de vue : en redéfinissant l'immunodéficience à une ou plusieurs infections comme la règle et

non comme l'exception, elle permet de comprendre que le fait d'être sensible à une famille d'infections (par exemple les infections mycobactériennes) dépend de la branche du système immunitaire qui est affectée : ce n'est pas par hasard si certains individus particulièrement sensibles aux infections mycobactériennes seront également particulièrement sensibles à l'encéphalite herpétique : c'est parce qu'une branche commune de la réponse immunitaire innée à ces agents infectieux est affectée par une mutation génétique.

Une autre difficulté de l'explication des maladies infectieuses que nous avons très brièvement évoquée en introduction et qu'Abel et Casanova soulèvent à plusieurs reprises dans leurs articles est de comprendre le lien qu'entretiennent la classe des maladies infectieuses avec d'autres classes de maladie, comme les cancers viro-induits, les maladies auto-immunes et les allergies. En effet, les maladies infectieuses peuvent être à l'origine du développement de cancer : c'est le cas par exemple de la famille des papillomavirus et en particulier des types HPV 16 et HPV 18. Dans 90% des cas, l'infection par HPV 16 et HPV 18 est asymptomatique et guérit spontanément, mais dans 10% des cas l'infection persiste et des cancers du col de l'utérus apparaissent dans un délai moyen de 10 ans après l'infection. Les maladies infectieuses sont par ailleurs soupçonnées d'agir comme déclencheurs d'un grand nombre de maladies auto-immunes, comme le diabète de type 1, la maladie de Crohn, la spondylarthrite ankylosante, etc. Enfin, des associations fréquentes entre infections virales et développement d'allergies chez l'enfant ont été mises en évidence. Intuitivement, on peut faire l'hypothèse que le lien entre infections, cancers, maladies auto-immunes et allergies n'est pas un hasard puisque toutes ces maladies touchent d'une façon ou d'une autre le système immunitaire : les maladies infectieuses produisent ainsi une réaction de défense du système immunitaire à un micro-organisme pathogène, les cancers peuvent être compris comme un échec du système immunitaire à contrôler le développement de cellules anarchiques, les maladies auto-immunes sont une réaction erronée du système immunitaire qui réagit contre les antigènes du soi et les allergies sont une réaction excessive du système immunitaire à un corps ou une substance étrangère. Néanmoins, en dehors de ce consensus de bon sens, et d'associations toujours plus nombreuses mises en évidence dans la littérature médicale contemporaine entre ces différentes classes de maladie, il n'existe pas aujourd'hui

d'explication plus développée de la comorbidité entre maladies infectieuses et ces autres classes de maladie. Que dit la théorie génétique des maladies infectieuses à ce sujet ?

Pour certaines maladies infectieuses, par exemple, les analyses de ségrégation et les études pangénomiques ont mis en évidence l'implication de variants alléliques dans le développement de la lèpre, dont le rôle aurait déjà été démontré dans la maladie de Crohn et dans l'infarctus du myocarde :

> « En plus de la maladie de Crohn et de l'infarctus du myocarde, une précédente revue sur la génétique humaine de la lèpre discute le lien génétique entre la lèpre et d'autres maladies communes comme l'asthme, l'arthrite psoriasique ou le diabète de type 1. Ainsi, la lèpre, maladie de la grande pauvreté, négligée bien plus qu'éliminée, se retrouve par le prisme de la génétique humaine au carrefour de nombreuses pathologies communes des pays les plus riches. Espérons que les fonds investis pour l'étude de ses dernières permettront par un effet boomerang mérité de mieux comprendre encore l'histoire naturelle de la lèpre avec l'espoir de proposer de nouveaux outils diagnostics et de développer de nouveaux moyens thérapeutiques pour aider les malades encore trop souvent stigmatisés. » (Gaschignard *et al*., 2013, p. 126)

Autrement dit, si la génétique humaine de la susceptibilité des maladies infectieuses permet de mettre ponctuellement en évidence l'existence de variants alléliques communs à des maladies infectieuses et à d'autres maladies, ces liens ne sont que des associations statistiques ponctuelles, sans hypothèse physiopathologique ou sans mécanisme général commun qui permettrait d'aboutir à une explication causale de la comorbidité entre maladies infectieuses, cancers, maladies auto-immunes et allergies. Un seul article de Casanova et Abel, publié en 2009, peut sembler faire figure d'exception. Dans cet article, les auteurs font l'hypothèse que la maladie de Crohn, une maladie inflammatoire chronique de l'intestin, généralement considérée comme une maladie auto-immune avec une composante héréditaire non négligeable, est en fait une immunodéficience primaire des macrophages, c'est à dire une nouvelle immunodéficience primaire (Casanova et Abel, 2009). Même si Casanova et Abel suggèrent une esquisse de mécanisme physiopathologique qui permettrait

d'expliquer le développement des symptômes de la maladie à partir de cette hypothèse, cet article n'est qu'un exemple d'une maladie auto-immune réinterprétée comme une nouvelle immunodéficience primaire : en aucun cas, les auteurs ne s'appuient sur un mécanisme général qui permettrait de subsumer plusieurs cas de comorbidité entre maladies infectieuses, cancers, maladies auto-immunes et allergies sous une même règle.

5.4. Conclusion du chapitre : de la théorie génétique des maladies infectieuses à une théorie génétique des maladies ?

Si nous avons souligné des insuffisances dans la théorie génétique des maladies infectieuses, on peut néanmoins conclure qu'elle constitue bien une théorie génétique régionale d'une classe de maladie qui donne une explication mécaniste de la variabilité interindividuelle aux maladies infectieuses et s'intègre en tant que théorie partielle dans un ensemble d'explications (théorie des germes, théorie microbiologique, théorie immunologique et théorie écologique) des maladies infectieuses. Une théorie génétique non génocentriste d'une classe de maladie est donc possible et peut même amener à des développements thérapeutiques que nous avons évoqués précédemment.

L'objectif de notre travail est de faire sens de la généticisation des maladies et de comprendre en quel sens cette généticisation des maladies constitue une théorie génétique des maladies, c'est-à-dire un ensemble de théories génétiques régionales (chaque classe de maladie ayant une théorie génétique correspondante), ou une théorie génétique de la maladie, c'est-à-dire une théorie susceptible de redéfinir le rôle commun des gènes dans toutes les maladies en modifiant éventuellement la façon dont nous classons habituellement les maladies. Comme il nous était impossible dans les limites de ce travail de rechercher l'existence d'une théorie génétique pour chaque classe de maladie, nous nous proposons donc de discuter à quel point les mécanismes génétiques qui fondent la théorie génétique des maladies infectieuses seraient potentiellement extrapolables à d'autres classes de maladie.

Des mécanismes dont les fondements conceptuels sont enracinés dans l'interactionnisme quantitatif, comme les concepts de « prédisposition mendélienne »,

de « prédisposition polygénique » et de façon générale, tout cadre qui discute de l'architecture allélique d'une classe de maladie, sont hautement extrapolables à d'autres classes de maladie, quand bien même l'architecture allélique ne serait pas strictement superposable d'une classe de maladie à une autre. Autrement dit, même si la proportion de variants rares / maladies communes ou si les modes de prédisposition ne sont pas strictement superposables, le fait que toutes les classes de maladie sont progressivement explorées avec les mêmes méthodes de génétique quantitative (héritabilité, analyse de ségrégation, analyse de liaison, associations pangénomiques et maintenant séquençage d'exome entier) permet probablement de garder un même cadre d'analyse pour toutes les classes de maladie, même si la définition de ces concepts issus de la génétique quantitative est sujette à discussion. De façon similaire, l'hypothèse d'une architecture allélique âge-dépendante peut probablement se transposer d'une classe de maladie à une autre. On peut donc imaginer sans difficulté un*e théorie génétique des maladies* en cours d'élaboration, constituée d'un ensemble de théories génétiques régionales, chacune de ces théories définissant pour une classe de maladie les mécanismes génétiques qui s'y appliquent et aboutissant à une description de l'architecture allélique d'une classe de maladie.

Si une telle théorie génétique des maladies aurait l'avantage d'offrir un cadre commun à l'analyse de la génétique humaine des maladies et a certainement un pouvoir heuristique fort à un moment de la génétique où les hypothèses sur l'architecture allélique des maladies se multiplient pour répondre au problème de l'héritabilité manquante, elle présente néanmoins plusieurs inconvénients. (1) Elle ne pourrait pas rendre compte des interactions gène x environnement qui sont au cœur d'une explication interactionniste co-constructionniste des maladies. C'est ce que nous avons examiné avec l'exemple de la théorie génétique des maladies infectieuses : si celle-ci peut s'accorder de principe avec la théorie des germes, la théorie microbiologique et la théorie immunologique, Casanova et Abel restent très flous quant aux liens qu'entretiennent la théorie génétique des maladies infectieuses avec la théorie écologique et plus généralement avec la variabilité des déterminants environnementaux des maladies infectieuse. (2) Une telle théorie ne permet pas non plus de rendre compte systématiquement des liens qui existent entre les classes de maladie. Le cas des maladies infectieuses, des cancers viro-induits, des maladies auto-

immunes et des allergies est emblématique de ce problème : la théorie génétique des maladies infectieuses n'a pas d'hypothèse générale bien construite qui permette de comprendre les relations de comorbidité ou de causalité entre ces différentes classes de maladie.

C'est pourquoi nous proposons de nous tourner à présent vers notre seconde hypothèse, consistant à interpréter la généticisation des maladies comme l'élaboration d'une théorie génétique de la maladie. Dans la mesure où une telle théorie n'est pas revendiquée de façon explicite dans la littérature biomédicale contemporaine, nous avons choisi de nous tourner vers des programmes de recherche comme la biologie et la médecine des systèmes, dont l'objectif n'est pas de comprendre les liens spécifiques entre un ensemble de gènes et une maladie-type ou entre un ensemble de gènes et une classe de maladie, mais d'explorer le rôle commun des gènes dans toutes les maladies.

Chapitre 6 : La médecine des réseaux, un cadre pour une théorie génétique de la maladie ?

L'objectif de cette troisième partie est de confronter nos deux hypothèses sur le statut épistémologique de la généticisation à la littérature biomédicale contemporaine. La première hypothèse est que la généticisation révèle l'élaboration d'une théorie génétique des maladies, c'est-à-dire d'un ensemble hétérogène de théories génétiques régionales. Le chapitre 5, consacré à la théorie génétique des maladies infectieuses nous a permis d'affirmer qu'il existe au moins un exemple de théorie génétique régionale non génocentriste dans la littérature biomédicale contemporaine, mais aussi de souligner les limites de cette théorie et de son extrapolation à d'autres classes de maladie. Le chapitre 6 consiste donc à tester notre seconde hypothèse, selon laquelle la généticisation révèle une théorie génétique de la maladie c'est-à-dire une explication du rôle des gènes dans toutes les maladies, qui pourrait avoir des conséquences à la fois sur la façon dont nous définissons la maladie et sur la façon dont nous classons les maladies.

Or, comme nous l'avons mentionné dans la présentation de cette troisième partie, il n'y a pas de théorie génétique explicite de la maladie dans la littérature biomédicale contemporaine. Nous avons donc choisi de nous tourner vers des programmes de recherche qui s'éloignent autant que possible d'explications individuelles de maladies données pour s'intéresser plutôt à la découverte des principes généraux d'organisation et de développement de la maladie, c'est-à-dire à la biologie et à la médecine des systèmes. Dans ce contexte de la médecine des systèmes, nous avons identifié la « médecine des réseaux », une approche développée récemment principalement par l'équipe d'Albert-Lázló Barabási[51], qui ne considère pas les maladies comme des entités distinctes et isolée mais comme des phénomènes physiopathologiques interconnectés unis par des principes communs de dysfonctionnement, qui peuvent être élucidés à différents niveaux d'organisation

[51] Albert-Lázló Barabási a commencé ses travaux en tant que physicien théorique à l'Université Notre Dame. Il est maintenant directeur du Centre pour la Recherche des Réseaux Complexes et membre associé du Centre de la biologie des systèmes cancéreux de l'Institut du cancer Dana Farber de l'Université Harvard. L'ensemble des travaux de son équipe ou des équipes associées au CCNR est disponible sur le site internet suivant : http://www.barabasilab.com/

(génétique, métabolique, etc.). Plus précisément, au niveau génétique, la médecine des réseaux est fondée sur l'hypothèse qu'il existe des « gènes humains de la maladie », distincts des autres gènes humains, qui ont des propriétés spécifiques que l'on peut analyser en utilisant un ensemble d'outils mathématiques, appelés la théorie des réseaux. L'objectif de ce chapitre est donc de discuter dans quelle mesure la médecine des réseaux est un cadre fécond pour l'émergence d'une théorie génétique de la maladie.

Nous commencerons par retracer rapidement les origines historiques et intellectuelles de la naissance de la médecine des réseaux. Puis nous présenterons le diseasome, l'un des outils principaux de la médecine des réseaux et qui est un graphique bipartite permettant de représenter l'ensemble des associations gènes-maladies sous forme de réseaux. C'est l'analyse topologique du diseasome qui permet d'identifier un ensemble de caractéristiques spécifiques des gènes humains de la maladie et de développer une définition modulaire de la maladie. Enfin, en reprenant le cadre conceptuel que nous avons développé dans la deuxième partie de cette thèse, nous expliciterons alors ce que la médecine des réseaux explique et comment elle l'explique. Plus particulièrement, nous montrerons en quoi la médecine des réseaux offre un cadre fécond pour le développement d'une théorie génétique de la maladie.

Enfin, nous reprendrons l'exemple des maladies infectieuses, en analysant comment la médecine des réseaux explique cette classe de maladie. Nous chercherons à déterminer à quel point cette explication est meilleure, compatible ou complémentaire de l'explication des maladies infectieuses développée dans la théorie génétique des maladies infectieuses, que nous avons analysée au chapitre 5.

6.1. Les origines de la médecine des réseaux

Si on peut déjà trouver des occurrences du terme « biologie des réseaux » (*network biology*) dès 2004 (Barabási et Oltvai, 2004), le terme « médecine des réseaux » apparaît à notre connaissance pour la première fois dans un article d'Albert-László Barabási de 2007, intitulé « Network Medicine – From Obesity to the « Diseasome » » et publié dans le *New England Journal of Medicine*. Barabási y commente les résultats d'une étude publiée dans la même édition du *NEJM*, qui s'est

intéressée aux liens sociaux entre les participants de la fameuse étude de Framingham sur le risque cardiovasculaire. L'étude conclut qu'appartenir au même réseau amical a un effet encore plus important sur le risque de devenir obèse que le fait de porter une seule copie de l'allèle FTO, qui augmente déjà le risque de devenir obèse de 30% par rapport à la population générale (Barabási, 2007). Barabási en conclut que l'étude des réseaux en médecine est appelée à connaître de nombreux développements.

Ce bref article expose quatre caractéristiques importantes de la médecine des réseaux: (1) la variété des applications que l'étude des réseaux peut avoir en médecine, aussi bien sur la connaissance des causes et de l'évolution des maladies que sur le développement pharmacologique de nouveaux traitements, (2) la variété des réseaux que l'on peut envisager d'inclure et d'étudier dans le cadre de la médecine et qui concernent aussi bien les réseaux de gènes, de protéines et de métabolites impliqués dans les maladies humaines, que les réseaux sociaux (voir Figure 33), (3) l'utilisation d'une même boîte à outils mathématiques pour analyser ces réseaux qui peuvent sembler si différents en apparence – la théorie des réseaux, que Barabási, dont la formation initiale est en physique théorique, a déjà développée dans le cadre de l'analyse du réseau internet, des réseaux de migration humaine ou des réseaux de composants biologiques, (4) la conviction qu'il faut sortir d'une vision dite « réductionniste » de la recherche biomédicale qui considère les maladies comme des événements distincts sans rapport les uns avec les autres pour passer à une perspective plus globale où les maladies sont comprises comme des événements interdépendants.

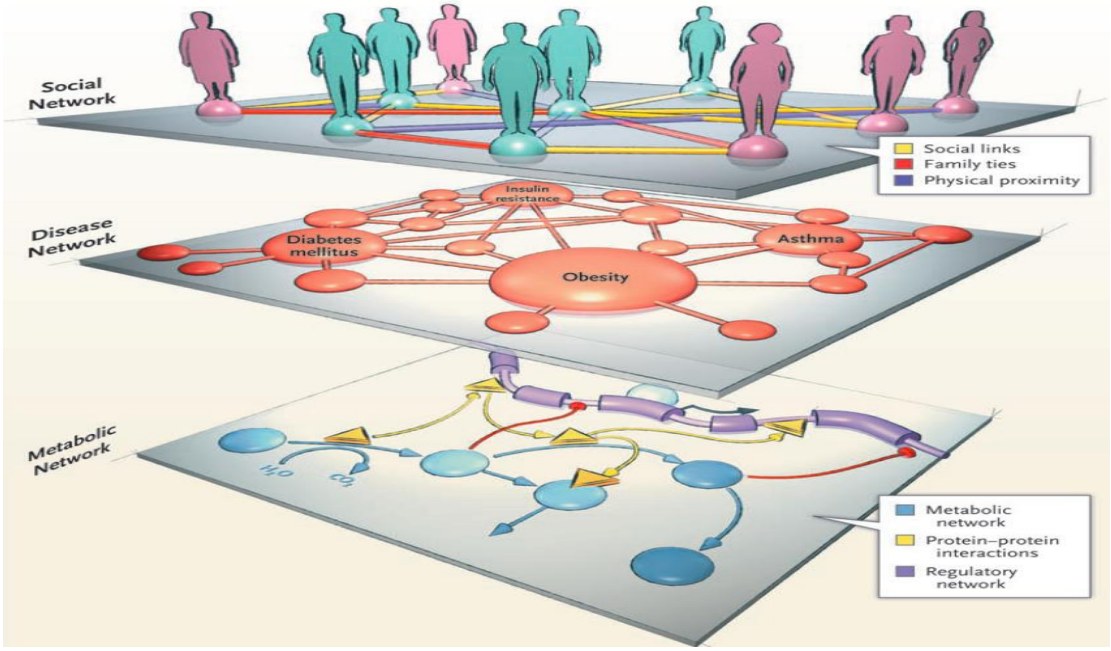


Figure 33 : Réseaux complexes qui sont directement pertinents pour la médecine des réseaux. (Barabási, 2007, p. 405). On distingue en particulier sur cette figure le réseau social qui étudie les interactions entre individus (proximité physique, liens familiaux ou sociaux), le réseau des maladies qui étudie les interactions entre des maladies considérées de prime abord comme différentes, et le réseau métabolique qui étudie les réactions métaboliques entre les différents composants moléculaires d'un individu.

Cette très rapide esquisse du rôle des réseaux en médecine sera développée de façon bien plus détaillée quatre ans plus tard dans un article publié en 2011 dans *Nature Reviews Genetics* (une revue de référence en génétique), qui signe véritablement l'acte de naissance de la « médecine des réseaux » (Barabási *et al.,* 2011). C'est d'ailleurs suite à la parution de cet article qu'un ensemble d'articles de référence sur la médecine des réseaux vont être publiés (Antman *et al.,* 2012 ; Arrell et Terzic, 2012 *;* Chan et Loscalzo, 2012 *;* Pawson et Linding, 2008 *;* Piro, 2012 ; Roukos, 2011 ; Silverman et Loscalzo, 2012, 2013 *;* Zanzoni *et al.,* 2009). Une recherche Pubmed du 17 mars 2014 retrouve ainsi 231 articles associés au terme « network medicine », dont 227 ont été publiés dans les 5 dernières années et dont 38 seulement contiennent l'expression « *network medicine* » dans le titre et/ou le résumé de l'article. Signalons également l'organisation récente de plusieurs symposiums internationaux sur le sujet : le symposium *« Integrative Network Biology: Network Medicine »* (« Biologie integrative des réseaux : la médecine des réseaux ») au Danemark au printemps 2012, le symposium *« Network Science in Biology and*

Medicine » (« La science des réseaux en biologie et en médecine ») à Harvard en juin 2012, l'ouverture d'un cours d'introduction à la médecine des réseaux à Harvard depuis 2013.

Nous insistons sur cette chronologie pour souligner le caractère extrêmement récent de la médecine des réseaux dont les objectifs, les concepts et les méthodes se sont essentiellement développés ces cinq dernières années. C'est bien évidemment une source de difficultés pour notre analyse en termes de recul historique. C'est pourquoi dans cette sous-partie, nous ne prétendons pas proposer une histoire exhaustive de la médecine des réseaux, mais seulement identifier et expliciter les principales idées fondatrices de ce tout nouveau champ disciplinaire. Trois points nous paraissent essentiels : (1) l'émergence de la médecine des réseaux s'est produite dans le contexte du développement récent d'une nouvelle méthode de recherche, la biologie et la médecine des systèmes, (2) la médecine des réseaux est centrée sur un nouveau concept, les gènes humains de la maladie, (3) elle utilise un nouvel outil d'analyse, la théorie des réseaux.

6.1.1. Un nouveau contexte : la biologie et la médecine des systèmes

La biologie des systèmes est un programme de recherche interdisciplinaire qui s'est considérablement développé ces quinze dernières années à la fois sur le plan institutionnel et académique (World Technology Evaluation Center, 2007). Ainsi, si on peut trouver l'expression « biologie des systèmes » dès la fin des années 1990 (Hood, 1998), c'est en 2000 que Leroy Hood et Hirotani Kitano, deux pionniers de cette discipline créent respectivement l'*Institute for Systems Biology* à Seattle et le *Systems Biology Institute* à Tokyo, tandis que le manifeste de la biologie des systèmes, « Systems biology – a brief overview », écrit par Hirotani Kitano paraît en 2002 dans *Science* (Kitano, 2002). On trouve différentes définitions de la biologie des systèmes (Braillard, 2009) et selon les auteurs, les fondements théoriques de ce programme de recherche peuvent être décelés à différents moments de la biologie du 20^{e} siècle (Carvunis *et al.*, 2009), mais certaines caractéristiques communes et récurrentes à ces différentes définitions peuvent être dégagées.

(1) L'objet de la biologie des systèmes est d'étudier les différentes interactions

entre les composants de systèmes biologiques et d'en tirer des principes généraux sur l'organisation et les propriétés de ces systèmes. Trois points méritent d'être soulignés : (a) la notion de système est une notion fluide en biologie : c'est l'expérimentateur qui définit les bornes d'un système biologique :

> « En biologie, une certaine confusion règne autour de cette définition car l'échelle d'étude d'un système peut lui valoir ou non l'appellation « système » selon l'opinion de celui qui le considère. En fait, il n'y a pas de limite théorique à la taille d'un système : libre à chacun de définir les bornes le délimitant de son environnement. Ainsi, un facteur de transcription régulant sa propre expression constitue un système, de la même façon que l'ensemble des molécules d'une cellule, l'ensemble des cellules d'un organisme, ou encore l'ensemble des individus d'une population. Différents systèmes peuvent être étudiés à partir des mêmes entités si l'on considère un type d'interaction plutôt qu'un autre. La population d'une ville peut être vue soit comme un ensemble d'individus partageant des relations économiques, des engagements matrimoniaux ou encore des maladies contagieuses, soit comme trois systèmes distincts. Un système peut aussi intégrer des composants hétérogènes, tel un écosystème comprenant à la fois des proies, des prédateurs et des ressources naturelles. Enfin, des composants et des interactions de natures différentes et traversant plusieurs échelles peuvent former un système unique. » (Carvunis *et al.,* 2009, p. 578 - 579)

(b) Puisque la définition des systèmes biologiques est très large, un des enjeux essentiels de la biologie des systèmes est l'intégration de données ou de propriétés de composants qui agissent à différents niveaux d'organisation et à différentes échelles d'un même système. (Lesne, 2009). (c) Malgré la diversité des systèmes biologiques étudiés, la biologie des systèmes entend mettre au jour un certain nombre de principes d'organisation communs en biologie comme la notion de robustesse ou de redondance fonctionnelle, sur lesquels nous reviendrons.

(2) La biologie des systèmes se présente souvent comme une approche « holiste » ou « anti-réductionniste » (Wolkenhauer et Green, 2013), qui considère l'ensemble des interactions entre les différents composants d'un système biologique,

par opposition à la biologie moléculaire qui s'intéresserait principalement à l'analyse isolée des composants du système. Pourtant, comme l'a montré Pierre-Alain Braillard (Braillard, 2009), cette opposition doit être relativisée et si la biologie des systèmes prétend et réussit à dépasser certaines limites de la biologie moléculaire, il y a davantage un « déplacement d'emphase » qu'une rupture entre ces deux traditions, dans la mesure où la biologie des systèmes utilise régulièrement les données de la biologie moléculaire pour comprendre les interactions entre les composants d'un système donné et combine (voir Figure 34) donc à la fois des approches dites *top-down* (du système aux composants) et *bottom-up* (des composants au système). Plusieurs scientifiques soutiennent donc le point de vue selon lequel la biologie des systèmes serait en continuité avec la biologie moléculaire (De Backer *et al.*, 2010), certains vont jusqu'à parler d'interdépendance méthodologique (Fang et Casadevall, 2011).

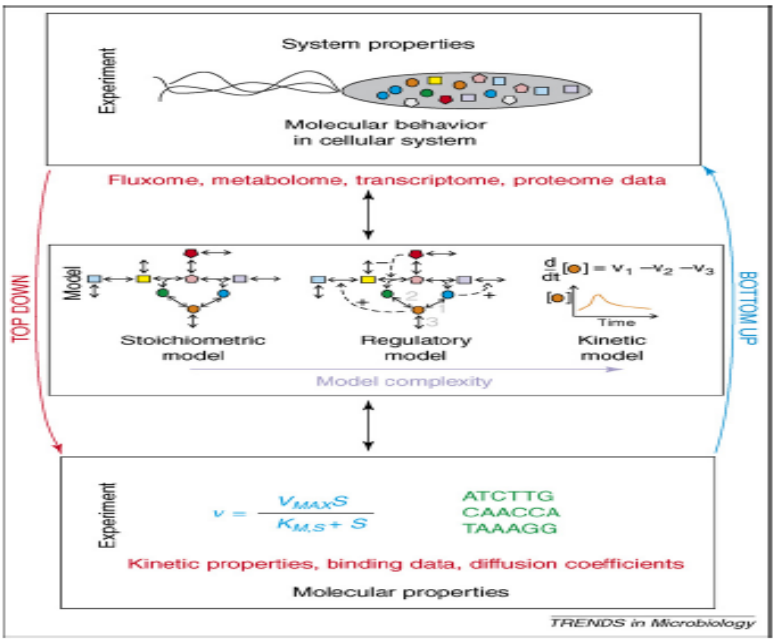


Figure 34 : Représentation de l'approche descendante ("top-down") et de l'approche ascendante ("bottom-up") dans la biologie des systèmes (Bruggeman et Westerhoff, 2007, p. 47). Les propriétés moléculaires (dérivées des expériences en biologie moléculaire, en biochimie et en bioinformatique) permettent de construire différents modèles (modèle stochiométrique, modèle de régulation, modèle cinétique, par exemple) mettant en relation différentes molécules (représentées par des formes géométriques de couleur et correspondant par exemple à des enzymes, des gènes, des facteurs de transcription, etc.). Dans l'approche ascendante, on utilise les propriétés moléculaires pour construire des modèles qui prédisent des propriétés du système, que l'on cherchera ensuite à valider expérimentalement, ce qui permet de raffiner le modèle (en comparant les écarts entre les propriétés prédites et le comportement du système). Inversement, l'approche descendante utilise des modèles pré-existants qui décrivent les données qui ont été mesurées dans le système, avec l'objectif d'identifier de nouvelles interactions, de nouveaux mécanismes et de nouvelles molécules.

(3) La biologie des systèmes dans sa forme contemporaine est fondée sur les bouleversements technologiques de l'ère génomique et post-génomique, en particulier sur l'avènement des technologies de séquençage et d'analyse de l'expression des gènes (puces ADN, puces ARN) à haut débit. Lorsque le terme de « génomique » est apparu en 1986 (Yadav, 2007), c'est déjà pour souligner les insuffisances de la génétique classique en tant qu'étude structurale de la séquence de chaque gène. Guttmacher et Collins définissent ainsi la génomique comme « l'étude non pas seulement de gènes isolés mais de leurs fonctions et de leurs interactions avec tous les gènes dans le génome » (Guttmacher et Collins, 2002, p. 1512). Pourtant, le temps de la génomique, qui est marqué par l'automatisation et l'amélioration des premières techniques de séquençage et qui aboutit la complétion du Projet Génome Humain, correspond encore à une vision structurale de la génétique, avec une focalisation sur le gène comme objet principal. Mais une fois le Projet Génome Humain terminé, lorsqu'à la surprise générale, seuls 25 000 gènes apparaissent nécessaires pour produire l'ensemble des protéines du corps humain, un nouveau cap est franchi et une nouvelle ère commence, celle de la « postgénomique ». La complexité biologique d'un organisme ne dépend pas seulement du nombre de gènes ou des interactions entre les gènes, mais également par des interactions entre les gènes et leurs produits, les protéines et des interactions entre les protéines et les autres composants cellulaires à l'échelle d'une cellule isolée, d'un tissu ou d'un organisme. La postgénomique correspond à une vision beaucoup plus fonctionnelle, contextuelle et dynamique du génome et une focalisation non plus tant sur l'ADN que sur les ARN transcrits et les protéines. Ce déplacement a été rendu possible grâce à l'apparition des biopuces d'épissage (puces ADN et ARN) à partir de 1995 (Lee et Hudson, 2000) : c'est ainsi que sont nées les nouvelles disciplines en « - omiques », comme la protéomique, dont l'objectif est d'identifier les protéines d'une population de cellules ou d'un tissu donné et qui permet d'analyser le rôle, la structure, la localisation et les interactions des protéines ou comme la transcriptomique, qui est l'étude de l'ensemble des ARN messagers transcrits à partir d'un génome dans un tissu particulier et dans un contexte physiologique ou pathologique donné. Une autre technique emblématique de la biologie des systèmes est la technique de double hybride qui permet la détection d'interactions physiques entre deux protéines. Admettons qu'on

veuille déterminer si la protéine A et la protéine B interagissent physiquement ensemble. Cette technique repose sur l'utilisation d'une protéine, un facteur de transcription, dont l'un des extrémités (*Binding Domain*, BD) peut se fixer à une séquence d'ADN et dont l'autre (*Activating Domain*, AD) est capable d'activer la transcription d'un gène rapporteur placé sous le contrôle de cette séquence d'ADN. Par génie génétique, on synthétise deux protéines hybrides : la première protéine hybride est le produit de la fusion de AD et de la protéine A, tandis que la seconde est une fusion de AD et de la protéine B. Si on détecte le produit du gène rapporteur (qui correspond la plupart du temps à une protéine fluorescente facilement détectable expérimentalement), cela signifie cela signifie que les domaines AD et BD ont été rapprochés, qu'un facteur de transcription fonctionnel a été reconstitué et donc que la protéine A et la protéine B interagissent. Cette technique, utilisée à grande échelle, symbolise l'emphase de l'ère postgénomique sur les études d'interactions entre les différents éléments du génome, mais aussi le développement des capacités d'analyse informatique qui permettent une analyse systémique de toutes les interactions protéine-protéine dans des organismes simples comme la levure par exemple.

(4) La biologie des systèmes s'est développée de façon concomitante de la bioinformatique, parce que les ambitions de la biologie des systèmes (étudier l'ensemble des interactions au sein d'un système biologique) aussi bien que l'explosion du nombre de données (les *Big Data* qui ont résulté des techniques de l'ère génomique et post-génomique) nécessitent la mise en place de bases de données (dont certaines bases de données publiques en ligne), l'uniformisation du format de ces données (d'où le développement des ontologies[52] en biologie), l'automatisation de techniques d'extraction (data-mining) et la mise en place de nouvelles formes de visualisation de ces données (Hagen, 2000 ; *Ram et al*., 2012).

(5) La biologie des systèmes fait appel à différentes techniques de modélisation mathématique pour traiter ces données et comprendre la nature des interactions

[52] Une ontologie est un ensemble structuré de termes et de concepts d'un domaine particulier, qui précise les relations entre ces termes et leurs propriétés. Chaque terme d'une ontologie doit posséder une définition pour être sûr de la signification qui y est associée.

entre les composants d'un système et les propriétés de ce système. Au sein de ces différentes techniques de modélisation mathématique, l'étude des réseaux d'interactions biologiques (réseaux des interactions gène-gène, réseaux des interactions gène-protéine, réseaux des interactions protéine-protéine,…), de la dynamique de ces réseaux et de la réaction de ces réseaux à des perturbations systématiques, a pris une grande importance, notamment avec l'essor de la science des réseaux que nous explorerons dans la section 6.1.3.

Quelles sont les applications de la biologie des systèmes à la médecine ? Autrement dit, comment passe-t-on de la biologie des systèmes à la médecine des systèmes ? Plusieurs auteurs, comme Leroy Hood (fondateur de l'*Institute for Systems Biology* à Seattle aux États-Unis), Charles Auffray (fondateur de l'*European Institute for Systems Biology and Medicine* à Lyon, en France) et Olaf Wolkenhauer (éditeur fondateur du journal *Systems Biology* en 2005 et professeur au centre de *Systems Biology and Bioinformatics* à Rostock en Allemagne), ont cherché à identifier dans une série d'articles à la fois la nature de la médecine des systèmes (Auffray *et al.,* 2009 ; Hood et Flores, 2012), la façon dont on peut passer de la biologie des systèmes à la médecine des systèmes (Clermont *et al., 2009* ; Vandamme *et al., 2013* ; Wolkenhauer *et al., 2013)* et les différents défis que la médecine des systèmes (Capobianco, 2012 ; Kolch et Kholodenko, 2013) doit surmonter pour avoir des applications biomédicales concrètes. Par-delà la diversité des définitions et la diversité des moyens proposés, plusieurs caractéristiques de la médecine des systèmes apparaissent déjà dans un rapport, publié à la suite de la troisième conférence internationale de biologie des systèmes en 2009 en Italie, intitulé « Systems Biology meets the clinic » :

> « Les participants de la conférence proposent que la médecine systémique soit définie comme l'application de la biologie des systèmes à la prévention, la compréhension, la modulation et la guérison des maladies du développement et des processus pathologiques de la santé humaine. Bien qu'il n'existe pas de limite claire entre la biologie des systèmes et la médecine des systèmes, on peut affirmer que la biologie des systèmes vise à une compréhension fondamentale des processus biologiques et ultimement, à la modélisation

> exhaustive des réseaux biologiques, alors que la médecine des systèmes souligne que l'objectif essentiel et la pertinence des modèles est translationnelle et vise des applications diagnostiques, prédictives et thérapeutiques. Par conséquent, les avancées de la médecine des systèmes doivent être évaluées à la fois à l'échelle médicale et à une échelle biologique plus basique, dans la mesure où la correspondance entre médecine et biologie est complexe. Certains modèles qui semblent fondamentalement biologiques peuvent avoir un impact médical important, alors que certains modèles moléculaires dont la complexité est impressionnante peuvent ne pas être immédiatement pertinents sur le plan médical. Alors que la biologie des systèmes s'est principalement focalisée jusqu'ici sur l'échelle moléculaire, la médecine des systèmes doit directement incorporer des informations cliniques à la mésoéchelle dans ces modèles ; en particulier, des variables cliniques classiques, des biomarqueurs et des données d'imagerie médicale. » (Clermont *et al.*, 2009, p. 88)

Ainsi (1) la démarcation entre biologie et médecine des systèmes n'est pas évidente, et bien évidemment la médecine des systèmes a vocation à s'appuyer sur les ressources de la biologie des systèmes, mais dans une perspective centrée sur les maladies. Cette perspective est à la fois théorique (mieux comprendre la physiopathologie d'une maladie à l'échelle d'une cellule, d'un tissu, d'un organe, d'un être humain, d'une population et à l'échelle d'une vie) mais aussi pratique (il s'agit de développer des marqueurs prédictifs de l'apparition et de l'évolution d'une maladie, ainsi que de nouveaux outils thérapeutiques – c'est pourquoi les participants du rapport parlent de « médecine translationnelle »). Dans un esprit similaire, Leroy Hood a développé l'idée que la médecine des systèmes devait conduire à un « changement de paradigme en médecine » et en particulier à l'avènement de la « médecine 4 P » : préventive, prédictive, personnalisée et participative (Galas et Hood, 2009 ; Hood, 2004 ; Hood et Flores, 2012).

(2) Ceci dit, la médecine des systèmes ne saurait être seulement une application pure et simple de la biologie des systèmes, non seulement parce que toute donnée ou tout modèle issus de la biologie des systèmes ne sont pas nécessairement pertinents dans le contexte de la recherche biomédicale ou de la clinique mais aussi

parce que la médecine des systèmes nécessite l'intégration de nouvelles données d'une échelle différente (la mésoéchelle ou échelle intermédiaire) comme les biomarqueurs, la description de phénotypes pathologiques, l'ensemble des informations sur le mode de vie susceptibles d'influencer l'évolution d'une pathologie, des scores cliniques permettant de suivre l'évolution d'une pathologie chez un individu au cours du temps, etc. C'est pourquoi Wolkenhauer, dans un article récent « The road from systems biology to systems medicine », propose une distinction entre une médecine des systèmes qui ne serait qu'une traduction ou qu'une application pure et simple des principes de la biologie des systèmes et une médecine des systèmes qui serait véritablement « intégrative » :

> « La biologie des systèmes développée pour les applications médicales ne représente pas le déploiement complet de la médecine des systèmes. Bien que des programmes biomédicaux ou même des programmes pertinents sur le plan clinique puissent être pris en compte par la recherche fondamentale, il y a une nécessité reconnue de « traduire » ces découvertes fondamentales dans la recherche clinique. Plutôt que de parler de « recherche translationnelle », nous défendons l'idée d'une « intégration des données cliniques et non cliniques ». L'idée n'est pas de traduire des concepts d'un domaine à un autre, comme on traduit les avancées dans les sciences matérielles en des produits améliorés pour le consommateur, mais d'intégrer des données à partir d'expériences in vitro, de modèles animaux et de données omiques de grandes échelles aux données cliniques sur les caractéristiques du corps, son fonctionnement, les biomarqueurs et les antécédents d'un patient. » (Wolkenhauer *et al.*, 2013, p. 503)

La médecine des systèmes apparaît donc comme un nouveau programme de recherche qui a commencé à se développer dans la deuxième moitié des années 2000 et qui est caractérisé par une compréhension de la maladie comme un système dynamique, qui résulte d'interactions entre des composants de différentes échelles (dans le temps et dans l'espace) et de différents niveaux d'organisation (de l'anomalie moléculaire jusqu'aux symptômes en passant par les états physiopathologiques

intermédiaires[53]). A ce stade, la médecine des systèmes reste largement programmatique : il faut créer et stocker ces nouvelles données, créer des ontologies médicales adaptées qui permettent leur extraction, leur manipulation et leur visualisation, intégrer des données qui ont des échelles très différentes, une dimension dynamique et qui proviennent de sources très variées (modèles animaux, expériences *in vitro*, modèles *in silico*) dans l'optique de développer une meilleure compréhension des principes généraux d'organisation de la maladie, ainsi que des outils préventifs, prédictifs et thérapeutiques qui soient autant que possible adaptés à l'histoire personnelle d'un patient donné.

6.1.2. Un nouveau concept : les « gènes humains de la maladie »

Parler des « gènes humains de la maladie » comme d'un concept novateur peut surprendre. Effectivement, à partir du moment où le concept de maladie génétique s'est éloigné du modèle de la maladie monogénique développé dans les années 1960 et où le concept de prédisposition polygénique a commencé à émerger, la chasse aux gènes a débuté et il semble que pour toute maladie il y ait des « gènes causant la maladie » (*disease-causing genes*) à découvrir. C'est d'ailleurs ce que nous avons montré dans le chapitre 3 en décrivant les différentes méthodes qui se sont mises en place pour détecter les variants alléliques responsables des maladies complexes et communes. Cette volonté de lier chaque maladie à un ou plusieurs gènes causant la maladie est également très claire dans les discours scientifiques qui entourent le Projet Génome Humain. Ainsi, dans le premier rapport du comité du NIH en 1998 (Wingerson, 1990), appelé « *Mapping our genes* », un chapitre entier est consacré aux applications du Projet Génome Humain en biologie et en médecine : la recherche des gènes humains de la maladie ne concerne pas seulement les gènes impliqués dans les maladies monogéniques mais aussi l'identification des maladies polygéniques. James Watson, le premier directeur du Projet Génome Humain affirme ainsi en 1990 :

> « Lorsque ils seront enfin interprétés, les messages génétiques encodés dans nos molécules d'ADN fourniront les réponses ultimes aux soubassements

[53] Les états physiopathologiques intermédiaires désignent des phénomènes infracliniques qui contribuent à l'explication des symptômes. Un exemple d'état physiopathologique intermédiaire est la réaction inflammatoire.

> chimiques de l'existence humaine. Ils ne nous aideront pas seulement à comprendre comment nous fonctionnons en tant qu'êtres humains sains, mais ils expliqueront également, au niveau chimique, le rôle des facteurs génétiques dans une multitude de maladies, telles que le cancer, la maladie d'Alzheimer et la schizophrénie, qui diminuent la vie quotidienne de tant de millions d'individus. ». (Watson, 1990, p. 44)

Mais c'est justement parce que l'objectif du Projet Génome Humain est de mettre en évidence et de décrire des variants alléliques pour chaque maladie individuelle qu'il y a eu très peu de tentatives pour unifier ces découvertes et pour trouver des principes communs au rôle des gènes dans les maladies – un point que Barton Childs et David Valle, tous deux professeurs à l'Université John Hopkins soulignent dans un article appelé « *Genetics, biology and disease* », en commentant leur décision de publier un nouveau journal, *The Annual Reviews of Genomics and Human Genetics* :

> « Aujourd'hui, à l'âge de la réduction, l'étude des détails moléculaires des maladies prédomine. C'est pourquoi on a accordé peu d'attention à l'idée selon laquelle il y aurait des principes qui gouvernent les origines et les caractéristiques de la Maladie en général plutôt que des maladies en particulier. » (Childs et Valle, 2000, p. 7)

La distinction entre la Maladie et les maladies est soulignée à plusieurs reprises dans le texte de Childs et Valle qui distinguent deux voies pour l'intégration de la génomique en médecine : soit on se contente d'une liste de variants alléliques reliés à des maladies individuelles, soit on cherche des principes généraux qui gouvernent « le côté génétique de la Maladie » :

> « Nous pouvons limiter notre usage du Projet Génome Humain en l'utilisant simplement comme un catalogue que nous appliquerions lorsque cela est nécessaire à chaque maladie, ou bien nous pouvons utiliser sa logique pour repenser le mode de fonctionnement et l'éducation médicale : nous pouvons nous inspirer de la perception du Vogel et Motulsky, qui considère la génétique comme la science fondamentale de la médecine. De toute évidence,

> nous ne pourrons pas répondre à des questions fondamentales sur la pathogénèse et le traitement, si ce n'est en prenant en compte les gènes et leurs produits. Nous ne pouvons pas avoir de vision globale de la médecine, si ce n'est dans le contexte des principes qui confèrent une unité et une continuité à l'apparente discontinuité de la santé et de la maladie. Ces principes promeuvent une cohérence en exerçant des forces centripètes sur la fragmentation centrifuge de la pensée réductionniste. » (Childs et Valle, 2000, p. 7)

Childs et Valle terminent cet article en mentionnant les travaux qu'ils publient l'année suivante avec Jimenez-Sanchez dans un article appelé *« Human Disease Genes »* dans *Nature*, et dans lequel apparaît à notre connaissance pour la première fois l'expression « gènes humains de la maladie », au sens où nous l'entendrons dans la suite de ce chapitre, c'est-à-dire non pas comme une liste de variants alléliques isolés sans rapport entre eux, mais comme un ensemble de gènes distincts de ceux qui ne causent pas de maladie (Jimenez-Sanchez et al., 2001). Étant donné l'importance fondatrice de cet article pour le concept de gènes humains de la maladie, il nous semble important de présenter son objectif principal et sa méthodologie.

L'objectif de cet article est d'observer l'ensemble des variants alléliques listés dans la base de données de la 7^{e} édition de « *Metabolic and Molecular Bases of inherited diseases* » qui contient 269 gènes, de rechercher sur la « *Morbid Map* » de OMIM les variants alléliques de ses gènes et leur localisation cytogénétique (923 variants) et de catégoriser chaque gène humain de la maladie en fonction de son produit protéique. De façon remarquable, rappelons qu'OMIM contenait à cette date essentiellement (à 97%) des gènes impliqués dans les maladies monogéniques, puisque les variants alléliques des maladies polygéniques ont été inclus tardivement dans OMIM comme nous le rappelions dans le chapitre 1. Les auteurs ont donc choisi d'exclure les 3% de gènes associés à des maladies purement somatiques (comme des formes non héréditaires de cancer) et les gènes du génome mitochondrial. Il ne s'agit donc pas encore de regarder l'ensemble des gènes humains de la maladie, mais au moins ceux qui sont le mieux répertoriés à cette date et qui concernent essentiellement les maladies mendéliennes monogéniques. Par ailleurs, pour chaque

gène, les auteurs ont procédé manuellement à une annotation fonctionnelle différente de celle choisie par le IHGSC (le consortium international pour le séquençage du génome humain). En effet, le système d'annotation du consortium repose sur un système d'annotation qui, certes, prend en compte le contexte biologique dans lequel le produit du gène est exprimé (activité biochimique, processus biologique, localisation cellulaire), mais qui ne prend pas en compte les caractéristiques de la présentation clinique de la maladie associée (âge de révélation de la maladie, mode d'hérédité, fréquence, sévérité, étendue du tissu touché, association à des malformations), que les auteurs de l'article ont justement extraites pour chaque maladie de la partie « littérature » de la base de données OMIM. Les auteurs tirent ainsi plusieurs conclusions :

(1) En comparant la classification fonctionnelle des produits protéiques correspondants à l'échantillon des gènes de la maladie étudiés, ils trouvent quatorze catégories, la catégorie la plus importante étant celle des gènes codant des enzymes (31%), et en deuxième position des gènes codant des protéines qui ont elles-mêmes un rôle modulateur de la fonction des protéines (protéines qui stabilisent, activent, répriment la fonction d'une seconde protéine (13,1%). Les douze autres catégories, qui représentent chacune moins de 10% du total, correspondent à des catégories appelées : protéines de fonction inconnue, protéines récepteurs, protéine facteurs de transcription, protéines de la matrice intracellulaire, protéines de la matrice extracellulaire, protéines des transporteurs transmembranaires, protéines des canaux, protéines de signalisation cellulaire, protéines pour des hormones, protéines de transporteur extracellulaire, immunoglobulines. La surreprésentation des enzymes pourrait être le fait d'un biais historique, puisque depuis Archibald Garrod, les erreurs innées du métabolisme ont fait l'objet d'une attention particulière. Cependant, plusieurs observations sur la répartition des gènes codant des enzymes chez la drosophile laissant à penser que dans les eucaryotes les plus évolués, la fraction des gènes codant des enzymes serait de 25 à 30% de l'ensemble du génome, c'est-à-dire relativement proche de l'échantillonnage.

(2) En comparant la corrélation entre classes fonctionnelles des produits protéiques et âge de début de la maladie : toutes les catégories fonctionnelles des protéines citées ci-dessus peuvent être impliquées dans des maladies à n'importe quel

âge de la vie, mais pas dans les mêmes proportions. Ainsi, les gènes codant des facteurs de transcription sont sur-représentés dans la catégorie « maladies *in utero* », ce qui reflèterait le rôle des facteurs de transcription dans l'orchestration du développement fœtal. Le corollaire de cette observation est d'ailleurs que les gènes encodant les facteurs de transcription représentent 30% des gènes liés à des maladies qui sont associées à des malformations. Les gènes codant des enzymes seraient responsables de 47% des maladies intervenant entre la naissance et l'âge de 1 an, alors que leur importance est moindre pour toutes les autres catégories d'âge[54]. Pour expliquer ce résultat, les auteurs supposent que les nourrissons utilisent le stock enzymatique maternel pendant les premiers mois de vie et que les premiers symptômes de ces maladies apparaissent lorsque ce stock maternel s'est épuisé.

(3) En corrélant les quatre classes fonctionnelles des produits protéiques les plus importantes (enzymes, modulateurs de la fonction de la protéine, récepteurs et facteurs de transcriptions) et les principales caractéristiques des maladies (fréquence, mode de transmission, âge de début, réduction de l'espérance de vie) : la première observation est que les maladies associées aux gènes de l'échantillon étudié sont rares ou très rares en fréquence, ce qui n'est pas étonnant puisque l'échantillon est composé essentiellement de gènes liés à des maladies monogéniques, dont on a déjà longuement rappelé qu'elles étaient peu fréquentes. Une seconde observation concerne le mode de transmission des gènes de l'échantillon : les gènes codant des enzymes sont majoritairement récessifs, alors que les gènes codant les modulateurs de protéines sont pour moitié dominants et pour moitié récessifs et qu'enfin les gènes codant des facteurs de transcription sont majoritairement dominants. Cette répartition du mode d'hérédité est assez cohérente avec les fonctions biologiques de chaque catégorie. Enfin, on observe différents pics d'âge de révélation de la maladie pour chaque catégorie fonctionnelle de protéine. Pour les facteurs de transcription, le pic est *in utero*, pour les enzymes, le pic a lieu dans la première année de vie, pour les récepteurs le pic a lieu entre 1 an et l'âge de la puberté et enfin, pour les modificateurs de la fonction des protéines, le pic a lieu à l'âge adulte. Plusieurs interprétations sont proposées par les auteurs : les gènes dont les produits protéiques sont des récepteurs

[54] Sauf pour les maladies qui surviennent après 50 ans, mais il est difficile de tirer des conclusions dans la mesure où il n'y a que 18 gènes impliqués dans cette catégorie.

seraient associées à des maladies débutant entre l'âge d'un an et la puberté, car c'est une période particulièrement intense de signalisation entre les cellules et les tissus. Quant aux gènes associés aux protéines modifiant la fonction d'autres protéines, ils seraient associés à des maladies débutant tardivement, car modifier la fonction d'une protéine ne signifie pas une perte de fonction complète : les individus porteurs de ce type de mutation deviendraient donc symptomatiques graduellement.

On peut discuter la méthodologie de l'article, le choix et la représentativité de cet échantillon nécessairement biaisé en faveur des maladies mendéliennes monogéniques, le choix des caractéristiques de la maladie prises en compte et de la méthode d'annotation fonctionnelle des produits protéiques. Mais ce qui nous paraît essentiel ici, c'est l'accent qui est mis sur la recherche de caractéristiques générales des gènes humains de la maladie et qui nous semble tout à fait dans la lignée de l'article précédent de Childs et Valle :

> « Nous croyons que l'approfondissement de nos connaissances sur les gènes associés aux maladies permettra aux chercheurs d'aborder des questions plus compliquées, comme la contribution aux maladies des gènes relativement au noyau biologique que partagent l'ensemble des espèces et relativement aux gènes qui encodent des protéines spécifiquement humaines, ou comme la façon dont les caractéristiques des séquences (comme la conservation ou le polymorphisme) sont liées aux caractéristiques des maladies et comme la façon dont la fonction des protéines est reliée aux résultats des thérapeutiques cliniques. » (Jimenez-Sanchez *et al.,* 2001, p. 854)

Cette caractérisation des gènes humains de la maladie par opposition aux gènes « non-humains » de la maladie et surtout par opposition aux gènes humains « qui ne causent pas des maladies » sera poursuivie par la même équipe, en association avec l'équipe d'Albert-Lázló Barabási, le grand promoteur de la théorie des réseaux et culminera dans deux articles publiés en 2007 et sur lesquels nous reviendrons dans la section 2 de ce chapitre : un article sur la classification des maladies à l'ère postgénomique et qui se présente comme « une approche dérivée des systèmes complexes appliquée à la pathobiologie humaine » (Loscalzo *et al.,* 2007) et un autre article qui propose justement d'analyser les interactions de ces gènes et de visualiser

leurs relations sous la forme de réseaux, créant ainsi le diseasome, le réseau des gènes humains de la maladie (Goh *et al.,* 2007).

6.1.3. Une nouvelle boîte à outils : la théorie des réseaux

La science des réseaux a principalement trois objectifs : identifier des propriétés statistiques pour caractériser la structure et le comportement de systèmes définis ou représentés comme des réseaux, créer des modèles de réseaux qui peuvent aider à comprendre la signification des propriétés des réseaux réels et prédire le comportement des systèmes composés ou définis comme des réseaux. La science des réseaux, aussi appelée « théorie des réseaux », dérive directement de la théorie des graphes, dont on attribue l'origine au mathématicien Euler, lorsqu'il s'est attaqué au « problème des sept ponts de Königsberg », consistant à trouver un parcours qui permette de revenir au point d'origine en passant une fois et une fois seulement par chacun des sept ponts de Königsberg. Euler serait le premier à avoir développé une représentation en graphe et un traitement mathématique de ce problème, chaque pont étant représenté comme un nœud relié par des arêtes (Figure 35).

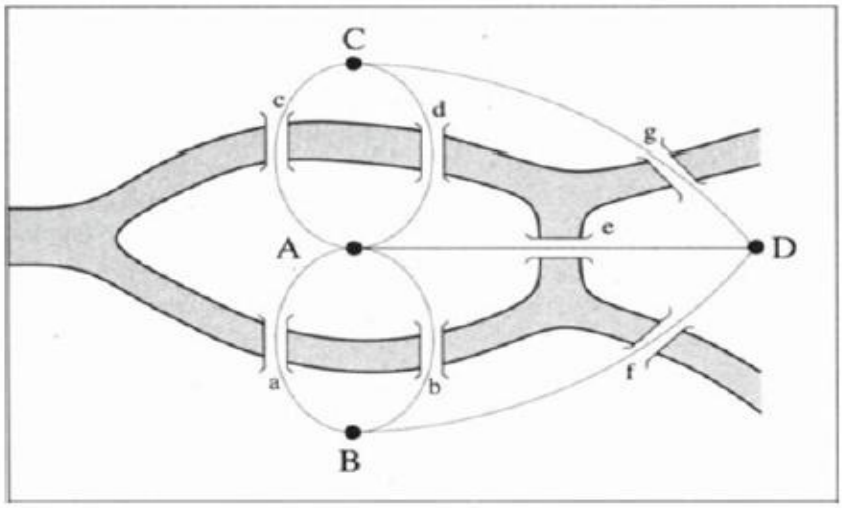


Figure 35 : Graphe surimposé sur une représentation des sept ponts de Königsberg (Baker, 2013, p. 691). La ville de Königsberg est construite autour de deux îles qui sont reliées par un pont. Pour relier les rives de la rivière à l'une ou l'autre des deux îles, il existe six autres ponts. Le problème consiste à déterminer s'il est possible de se promener dans Königsberg (quel que soit le point de départ) en passant une et une seule fois par chaque pont, pour revenir finalement à son point de départ. Euler a démontré qu'une telle promenade n'existait pas en introduisant la notion de « graphes eulériens ».

On distingue ainsi deux sortes de graphes : dans les graphes orientés (Figure 36B), les nœuds sont reliés par des arêtes avec une origine et une extrémité, tandis que dans les graphes non orientés (Figure 36A), les nœuds sont reliés par des arêtes non fléchées. L'intérêt des graphes orientés est de décrire le sens d'une interaction, par exemple de rendre compte de la direction des réactions dans un réseau métabolique. On appelle N le nombre total de nœuds dans un réseaux. Par exemple, dans la Figure 36A et dans la Figure 36B, N = 6. Par convention, on appelle k le degré d'un nœud ou sa connectivité : c'est le nombre de nœuds auxquels est connecté un nœud donné. Par exemple, dans la figure Figure 36A, k(1) = 2 et k(6) = 3. On appelle P(k) la probabilité qu'un nœud donné soit connecté exactement à k nœuds.

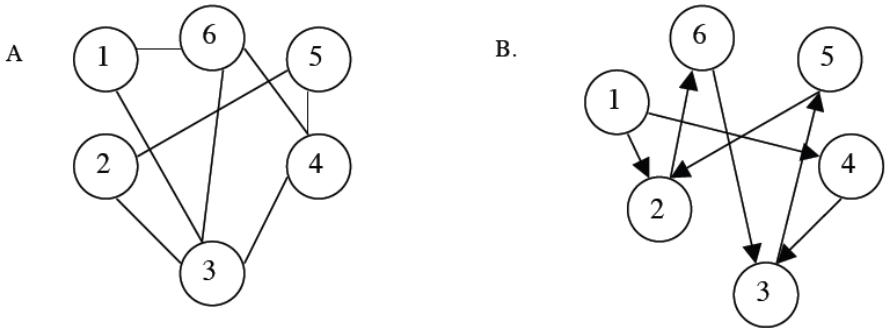


Figure 36 : Représentation d'un graphe non orienté (A) et d'un graphe orienté (B) (Braillard, 2009, p. 242) Dans un graphe orienté, les arêtes n'ont pas de direction, alors que dans un graphe orienté, le sens de l'interaction est précisé.

La propriété de connectivité k permet de distinguer deux types de réseaux dont la structure diffère fondamentalement : les réseaux aléatoires, dont les propriétés ont été explorées par les mathématiciens hongrois Paul Erdos et Alfred Renyi dans les années 1960 et les réseaux invariants d'échelle, dont les propriétés ont été explorées par Barabási à partir de 1999 (voir Figure 37).

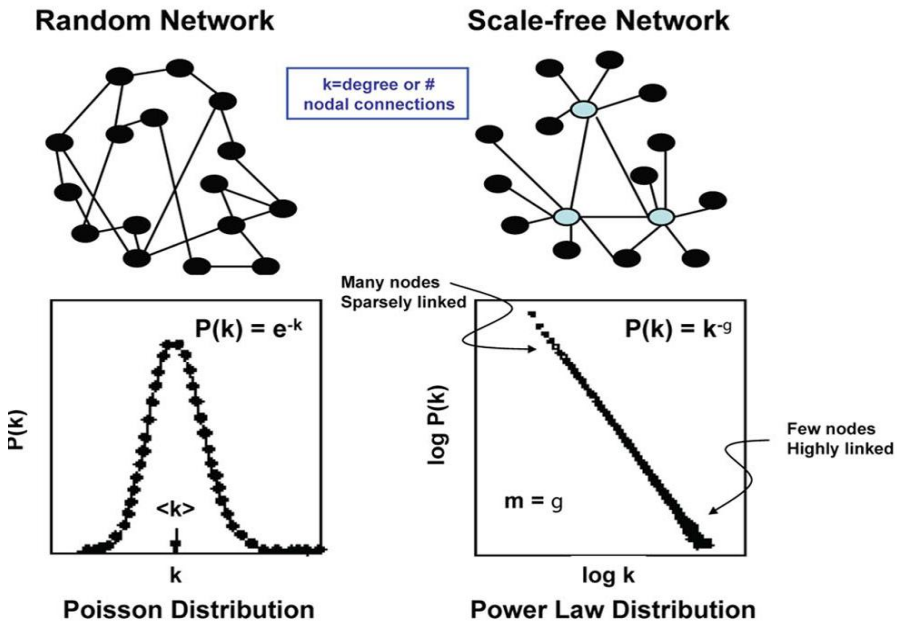


Figure 37 : Différences entre un réseau aléatoire et un réseau invariant d'échelle (Chan et Loscalzo, 2012, p. 364). A gauche, un réseau aléatoire est caractérisé par une distribution de la connectivité qui suit la loi de Poisson : le degré moyen des nœuds est peu élevé, la majorité des nœuds ont le même nombre (faible) d'interactions. A droite, un réseau invariant d'échelle est caractérisé par une distribution de la connectivité qui suit la loi de puissance : quelques nœuds (représentés en bleu clair) sont très connectés et forment des *hubs*, tandis que la majorité des nœuds sont faiblement connectés au reste du réseau.

Dans un réseau aléatoire, la probabilité qu'un nœud soit connecté à k autres nœuds suit une loi de Poisson, qui s'exprime par la formule $P(k) = e^{-k}$. Pour le dire plus simplement, tous les nœuds ont une connectivité équivalente ou proche, c'est-à-dire que tous les nœuds ont environ la même probabilité d'être connecté à k autres nœuds. Dans les réseaux invariants d'échelle, au contraire, un petit nombre de nœuds sont extrêmement connectés (on les appelle des « *hubs* » ou « carrefours[55] ») tandis que la majorité des nœuds sont très faiblement connectés. Dans les réseaux invariants d'échelle, la distribution des nœuds suit donc une loi de puissance, exprimée par la formule $P(k) = k^{-\gamma}$ où γ est l'exposant de degré ; la distribution de la connectivité P(k) est donc proportionnelle à γ. L'exposant de degré γ est un paramètre important qui détermine un grand nombre de propriétés du réseau étudié. Plus la valeur de γ est petite et plus le rôle des *hubs* est important dans le réseau. Ainsi, pour $\gamma > 3$, il y a peu

[55] La traduction « carrefour » n'étant jamais employée dans la littérature scientifique française au sujet de la science des réseaux, nous conserverons dans la suite de cet exposé le terme « *hubs* ».

de différences entre les *hubs* et les autres nœuds, c'est-à-dire que la plupart des propriétés topologiques spécifiques aux réseaux invariants d'échelle sont absentes et que le réseau se comporte de façon très similaire à un réseau aléatoire. En revanche, lorsque γ est compris entre 2 et 3, il y a une hiérarchie entre les *hubs*, le *hub* le plus connecté étant en contact avec une petite fraction de l'ensemble N des nœuds. Dans cette classe spécifique de réseaux invariants d'échelle, on trouve des réseaux hiérarchiques (que nous définissions plus bas). Enfin, pour $\gamma = 2$, un réseau « en étoile » (« *hub and spoke network* ») apparaît, le *hub* le plus large étant en contact avec une très large fraction de l'ensemble N des nœuds du réseau.

Notons, comme le remarque Barabási, que le terme « réseau invariant d'échelle », parfois aussi traduit « réseau sans échelle » est ambigu :

> « Les réseaux avec une loi de puissance sont appelés réseaux invariants d'échelle, un nom qui tire son origine de la littérature en physique statistique. Il indique l'absence de nœud typique dans le réseau (un nœud qui pourrait être utilisé pour caractériser les autres nœuds). Ceci contraste fortement avec les réseaux aléatoires, pour qui les degrés de tous les nœuds sont voisins du degré moyen, qui pourrait être considéré comme typique. Cependant les réseaux sans échelle pourraient aussi être appelés réseaux riches en échelles, puisque leur principale caractéristique est la coexistence de nœuds avec des degrés (ou échelles) extrêmement variés, allant de nœuds avec un ou deux liens à des *hubs* majeurs. » (Barabási et Oltvai, 2004, p. 104)

Autrement dit, les réseaux sont dits invariants d'échelle ou sans échelle, parce qu'ils n'ont pas d'échelle intrinsèque ou de seuil permettant de distinguer les nœuds des *hubs*. C'est pourquoi la définition des *hubs* ne peut être qu'arbitraire et par convention, on appelle *hubs* les 20% de nœuds avec le plus haut degré d'un réseau. En même temps, il existe une variété d'échelles qui coexistent et pour les réseaux dont l'exposant de degré est entre 2 et 3, cette coexistence de degrés variés crée une forme de réseaux hiérarchiques, qui sont des réseaux dans lesquels certains sous-graphes (un

groupe de nœuds liés les uns aux autres qui forme un sous-réseau au sein d'un réseau) reviennent de façon récurrente et sont appelés « motifs ».

Trois paramètres peuvent être utilisés pour évaluer le degré de modularité, c'est-à-dire la capacité d'un réseau à contenir des motifs qui fonctionnent comme des sous-unités en partie indépendantes du réseau général. Le coefficient de *clustering* local (*local clustering coefficient*), généralement noté C_v, est défini comme la probabilité moyenne que deux voisins d'un nœud soient adjacents et mesure le degré de *clustering* d'un nœud donné au sein d'un réseau. Autrement dit, sachant que A est connecté à B et que B est connecté à C, le coefficient de *clustering* local C_B permet d'estimer la probabilité que A et C soient également connectés. Le coefficient moyen de *clustering* (*average clustering coefficient*), généralement noté <C>, est défini comme la moyenne des C_v pour tout nœud v du réseau et mesure la tendance générale du réseau à générer des motifs. Enfin, la fonction C_k est le coefficient moyen d'agglomération de tous les nœuds ayant k liens. La propriété C et la propriété C_k (voir Figure 38) sont différentes pour des réseaux aléatoires, des réseaux invariants d'échelle et des réseaux hiérarchiques. En particulier, pour les réseaux invariants d'échelle, le coefficient moyen de *clustering* <C> est dépendant de la taille du réseau (c'est-à-dire du nombre total de nœuds constituant le réseau) : plus le nombre de nœuds est important, plus le coefficient moyen de *clustering* diminue. Dans les réseaux hiérarchiques, ce coefficient moyen de *clustering* est au contraire indépendant de la taille du réseau. Enfin, alors que le coefficient de *clustering* C_k est à peu près constant dans un réseau sans échelle (c'est-à-dire que dans un réseau invariant d'échelle, le coefficient d'agglomération est à peu près le même pour un nœud, quel que soit son nombre de liens), C_k suit une loi des échelles dans les réseaux hiérarchiques, exprimée comme $C_k \sim k^{-1}$ pour un nœud de k liens : c'est la signature d'une structure modulaire.

Un autre paramètre caractéristique de la structure d'un réseau est la longueur de chemin l qui mesure le nombre d'arêtes nécessaires pour connecter deux nœuds donnés. Comme il existe parfois plusieurs itinéraires possibles (et donc plusieurs longueurs de chemins différentes) pour connecter un nœud à un autre, on mesure en priorité le plus court chemin, c'est-à-dire le chemin comportant le moins d'arêtes pour connecter un nœud à un autre. Dans les réseaux aléatoires, comme dans les réseaux

invariants d'échelle, on retrouve une propriété commune appelée « propriété *small world* » : dans un réseau, deux nœuds peuvent être connectés par un petit nombre d'arêtes. Les réseaux invariants d'échelle sont même dits des « *ultra-small-worlds* » : par exemple, dans les réseaux métaboliques, trois à quatre réactions suffisent en moyenne à relier la plupart des métabolites entre eux. Une autre propriété qui peut être évaluée à partir de la longueur de chemin est la « *betweenness centrality* » d'un nœud. La *betweenness centrality* d'un nœud est la mesure du nombre de plus courts chemins qui passent par un nœud donné : plus la *beetweeness centrality* d'un nœud est importante, plus celui-ci tend à être un goulot d'étranglement dans le réseau.

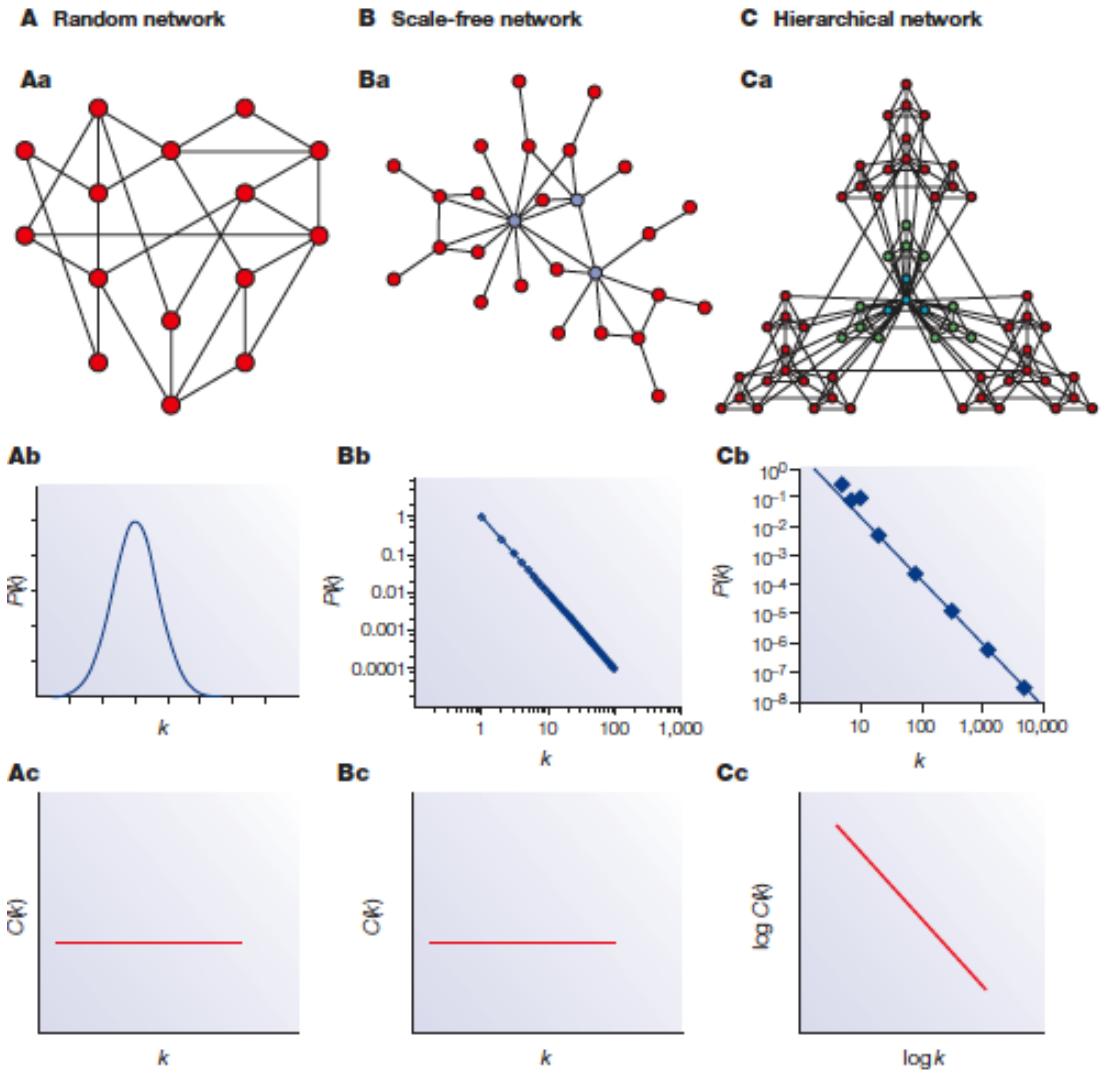


Figure 38 : Représentation des propriétés des réseaux aléatoires, invariants d'échelle et hiérarchiques (Barabási et Oltvai, 2004, p. 105). En haut, est représentée la forme des différents réseaux dans l'espace. Au milieu, est représentée la distribution de la connectivité dans les différents types de réseaux. En bas, est représenté le coefficient de *clustering* de chaque type de réseau.

L'intérêt de l'approche développée par Barabási (Barabási, 1999) est d'avoir démontré qu'un grand nombre de réseaux réels, aussi variés que les réseaux sociaux de communication téléphonique (Onnela et al., 2007), les réseaux écologiques (Montoya *et al.,* 2006) ou les réseaux internet (Yook *et al.,* 2002), se comportaient comme des réseaux invariants d'échelle et non comme des réseaux aléatoires. En particulier, les réseaux biologiques se comportent également comme des réseaux invariants d'échelle, ce qui justifie la naissance de la *network biology* (biologie des réseaux), qui est l'étude de la structure et des fonctions des réseaux biologiques. Par exemple, dans un réseau métabolique, les métabolites sont représentés par des nœuds et les réactions biochimiques par des arêtes orientées. L'analyse du métabolisme de 43 organismes différents appartenant au royaume des archées, des bactéries et des eucaryotes, indique que le métabolisme cellulaire suit une topologie de réseau invariant d'échelle, dans laquelle la plupart des substrats métaboliques participent seulement à une ou deux réactions, tandis qu'un petit nombre de métabolites participent à plusieurs douzaines de réactions et fonctionnent comme des *hubs* métaboliques (Jeong *et al.,* 2000*)*. Un autre exemple de réseau biologique est le réseau d'interaction protéine-protéine de la levure *Saccharomyces cerevisiae*, reconstruit à partir de données obtenues par la technique du double hybride. Comme on peut le voir dans la Figure 39, si la plupart des protéines participent seulement à un petit nombre d'interactions, un petit nombre de protéines sont au contraire impliquées dans un grand nombre d'interactions et sont des protéines *hubs*.

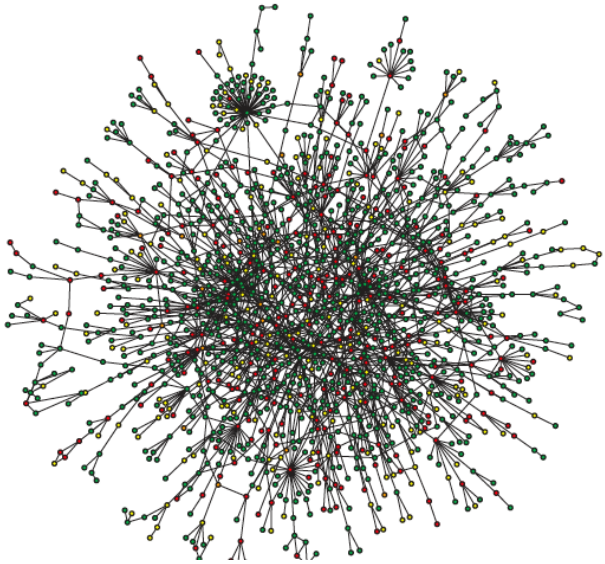


Figure 39 : Réseau d'interactions protéine-protéine de la levure *(Jeong et al.,* 2001). Ce réseau a été reconstruit à partir de la mise en évidence d'interactions protéine-protéine chez la levure par la technique du double hybride. C'est l'un des premiers exemples de réseau invariant d'échelle en biologie.

Comment expliquer que des réseaux réels aussi différents aient une structure de réseaux invariants d'échelle ? Barabási a proposé deux mécanismes pour expliquer la formation des réseaux invariants d'échelle, celui de la croissance et de l'attachement préférentiel (Barabási, 1999). Le mécanisme de croissance suppose que le réseau grandit par addition itérative de nouveaux nœuds. Le mécanisme d'attachement préférentiel est parfois résumée par l'adage « *the rich get richer* » (les riches s'enrichissent) et signifie qu'un nouveau nœud se connectera préférentiellement au nœud le plus connecté du réseau (Figure 40).

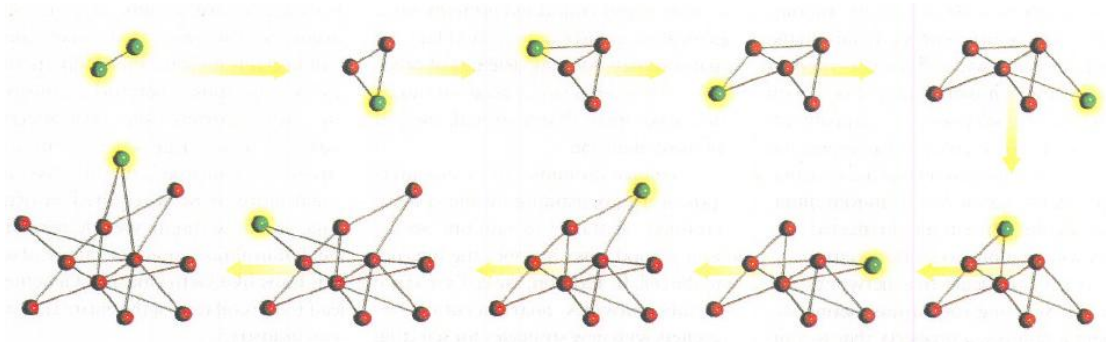


Figure 40 : Naissance d'un réseau invariant d'échelle par croissance et attachement préférentiel (Barabási et Bonabeau, 2003, p. 55). Selon cette hypothèse, chaque nouveau nœud du réseau a tendance à se lier aux nœuds les plus connectés, illustrant la règle selon lesquelles « les riches s'enrichissent ».

La crédibilité de ces mécanismes a été évaluée de plusieurs façons. La première est un processus de modélisation *in silico* (Barabási, 1999) qui en combinant ces deux mécanismes de façon itérative permet de créer un réseau invariant d'échelle (voir Figure 40) qui possède bien les propriétés d'usage (la distribution des degrés suit bien une loi de puissance dans le réseau obtenu). Par ailleurs, un tel modèle de croissance implique que les nœuds les plus connectés sont les nœuds les plus vieux d'un point de vue évolutionnaire, une prédiction qui a été confirmée par l'analyse des réseaux métaboliques où les coenzymes A, NAD et GTP, qui sont impliquées dans les chemins métaboliques les plus anciens comme la glycolyse, sont les nœuds les plus connectés (Wagner et Fell, 2001). Plusieurs études de réseaux d'interaction protéine-protéine dans des génomes d'espèces différentes ont également montré qu'en moyenne, les protéines les plus connectées étaient les plus vieilles d'un point de vue évolutionnaire (Wagner, 2003). Enfin, cette hypothèse pourrait être compatible avec une hypothèse biologique de duplication génétique au cours de l'évolution (voir Figure 41). Les mécanismes de formation des réseaux biologiques restent cependant sujets à discussion (pour une revue récente de ces débats, voir Zhao et Mooney, 2012 ; voir également, Barabási, 2012).

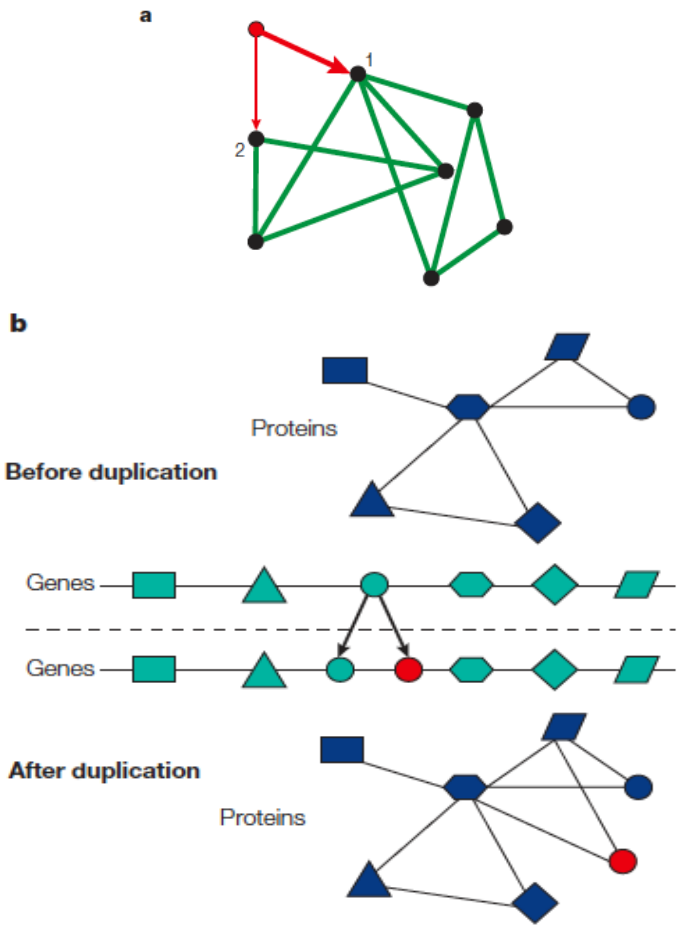


Figure 41 : Origine évolutionnaire et mécanisme de développement d'un réseau invariant d'échelle : (Barabási et Oltvai, 2004, p. 107) Figure A : L'attachement préférentiel signifie que dans ce réseau, le nœud rouge a deux fois plus de chances de se connecter au nœud 1 qu'au nœud 2, puisque le degré du nœud 1 est deux fois supérieur au nœud 2 (k_1=4 et k_2=2). Figure B : représentation d'un petit réseau d'interactions protéine-protéine (les protéines sont représentées en bleu) et des gènes qui les encodent (représentées en vert). Quand une cellule se divise, occasionnellement, un ou plusieurs gènes sont copiés sans modification dans le génome de leur descendance (cercles rouges et verts), ce qui se traduit par la croissance du réseau d'interaction protéine-protéine puisqu'il y a un gène supplémentaire qui encode une nouvelle protéine (cercle rouge). Finalement, les protéines qui interagissaient avec la protéine qui a été dupliquée gagne une nouvelle interaction avec la nouvelle protéine. C'est pourquoi les protéines avec un large nombre d'interactions gagnent des liens plus souvent : du fait de leur grand nombre de liens, il est plus probable qu'elles interagissent avec la protéine qui a été dupliquée. C'est ce mécanisme qui génère l'attachement préférentiel dans les réseaux cellulaires. En effet, dans cet exemple, peu importe quel gène est dupliqué, c'est la protéine *hub* qui gagne une interaction. Au contraire, la protéine représentée par un carré, qui n'a qu'un seul lien, gagne un nouveau lien si et seulement si la protéine *hub* est dupliquée.

L'invariance d'échelle des réseaux biologiques a une conséquence importante en terme de robustesse. Ce terme fait grandement débat en biologie des systèmes (pour une introduction à cette notion dans le contexte de la biologie des systèmes, voir Kitano, 2004) mais on peut le définir comme la capacité d'un système biologique à

résister à un ensemble de perturbations. Reste encore à définir comment on établit les bornes du système biologique en question, ce qu'on entend par capacité à résister (stabilité de la structure, capacité à assurer une fonction, capacité à adapter une réponse proportionnée, survie du système, etc.), et comment on définit l'ensemble des perturbations (Lesne, 2008). Dans le contexte de la théorie des réseaux, on distingue pour décrire la capacité de maintien de la structure du réseau face à deux types de perturbations : des perturbations aléatoires (attaques au hasard de plusieurs nœuds du réseau) et des perturbations ciblées (attaques de nœuds spécifiques). Comme on peut le voir dans la Figure 42, dans les réseaux aléatoires, quelques attaques aléatoires suffisent à désorganiser complètement le réseau. En revanche, pour les réseaux invariants d'échelle, des attaques aléatoires (par exemple, des mutations aléatoires induites par un carcinogène comme le tabac ou les UV) ont peu de conséquences sur la structure générale du réseau : au contraire, des attaques ciblées sur les protéines *hubs* du réseau sont déterminantes.

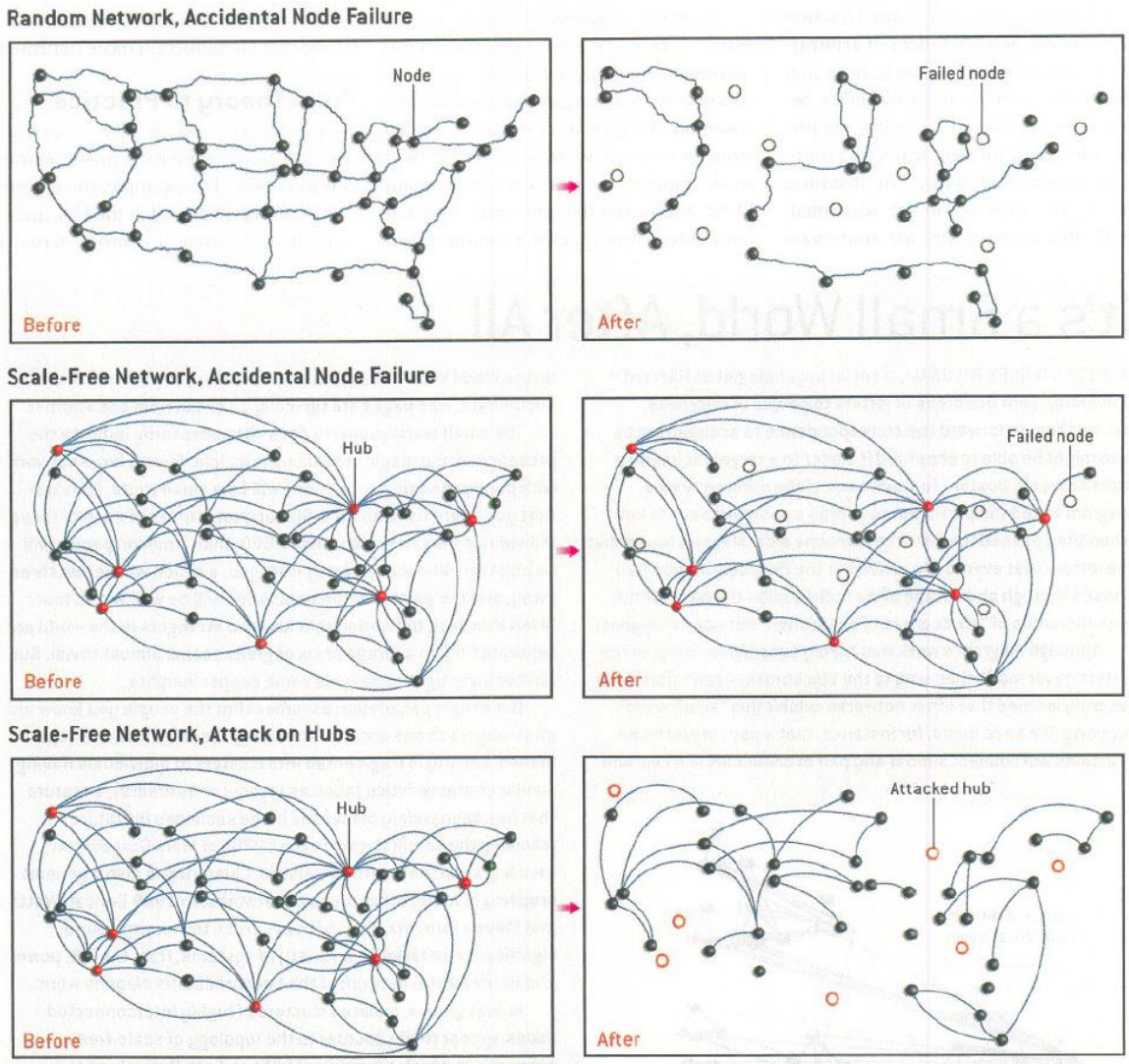


Figure 42 : Différents types de robustesse à différents types de perturbations pour les réseaux invariants d'échelle et les réseaux invariants aléatoires (Barabási et Bonabeau, 2003, p. 57).

Si le développement de cette nouvelle « science des réseaux » a suscité un vif enthousiasme, il a également inspiré de fortes réserves. L'un des exemples les plus provocateurs et emblématiques de ces critiques de la théorie des réseaux est un article intitulé « Revisiting scale-free network » (Keller, 2005) où Evelyn Fox Keller affirme a) que la loi de puissance n'est pas un principe nouveau et que Simon avait déjà mis en évidence ce type de structure b) que la loi de puissance peut donner lieu à des architectures de réseau très différentes c) que l'attachement préférentiel et la croissance ne sont qu'un mécanisme parmi d'autres possibles pour parvenir à un réseau invariant et qu'il n'est pas possible sans avoir d'information sur les contraintes qui ont pesé sur la formation d'un réseau donné de déterminer lequel de ces

mécanismes est responsable de sa structure c) qu'il est nécessaire de prendre en compte la spécificité des réseaux (réseau biologique, réseau social) pour véritablement comprendre leur organisation et tirer de la théorie des réseaux des informations pertinentes.

Nous souscrivons à ces critiques, mais les comprenons davantage comme une mise en garde contre l'omniprésence de la théorie des réseaux dans la littérature scientifique contemporaine et comme un appel à la prudence lorsqu'on analyse un réseau donné que comme une remise en cause des outils mathématiques que constitue la science des réseaux. Evelyn Fox Keller critique avant tout la prétention de la physique (dont vient initialement Barabási) à proposer une « loi universelle » qui serait valable pour l'organisation de tous les réseaux. On pourrait cependant objecter (réponse au point c) que la plupart des partisans de biologie des systèmes ont bien conscience que la structure du réseau n'est qu'un paramètre du système et n'explique pas à elle seule le comportement de ce dernier, en particulier lorsqu'il s'agit d'un réseau biologique. Les paramètres dynamiques, la cinétique des interactions (qui peuvent être stables ou transitoires), et la fonction moléculaire de chaque élément sont pris en compte dans un certain nombre de travaux de biologie systémique. Par exemple, le groupe de Marc Vidal à Harvard a intégré le profil d'expression des ARN transcrits de la levure à l'analyse du réseau d'interaction protéine-protéine de la levure (Han *et al*., 2004). Cette intégration d'informations dynamiques et fonctionnelles a permis d'affiner le concept topologique de protéines *hubs* et d'introduire une distinction conceptuelle entre deux types de protéines *hubs* : les « *date hubs* » et les « *party hubs* ». Les *party hubs* interagissent avec la plupart de leurs partenaires simultanément, tandis que les *date hubs* se lient à différents partenaires à différents endroits et à différents moments. Les premiers agiraient préférentiellement à l'intérieur de modules fonctionnels, tandis que les seconds connecteraient plutôt les modules fonctionnels entre eux.

Ces critiques permettent aussi de rappeler un point fondamental : les réseaux invariants d'échelle comme les réseaux aléatoires ne sont que des modèles utilisés pour comprendre ou reconstruire un ensemble de données disparates sur un système biologique réel donné. C'est pourquoi nous avons parlé dans le titre de cette section de la théorie des réseaux comme d'une « boîte à outils » plutôt que d'une « science ».

C'est également pourquoi lorsqu'on analyse une publication scientifique sur les réseaux biologiques, il est essentiel de discuter comment les données qui sont analysées avec le modèle des réseaux invariants d'échelle ont été obtenues (expérimentalement ou *in silico*, base de données utilisée, etc.).

Une autre critique est celle de Pierre-Alain Braillard, qui identifie deux grands courants dans la biologie des réseaux aujourd'hui : d'un côté, des travaux comme ceux de Barabási, qui relèvent d'une approche *top-down* et qui visent à utiliser des principes généraux sur la structure des réseaux pour comprendre les interactions entre leurs parties, et de l'autre, des travaux comme ceux d'Alon Uri qui cherchent à identifier les motifs que nous avons évoqués précédemment et qui sont compris comme les composants élémentaires des réseaux pour reconstruire par une méthode *bottom-up* l'organisation d'un système biologique donné (Braillard, 2009, 247-248). Braillard examine davantage en détail les travaux d'Uri, arguant que cette approche permet une meilleure intégration des données moléculaires spécifiques des réseaux biologiques et une meilleure compréhension non seulement de la structure des composants du réseau, mais surtout de leur fonction. Barabási lui-même reconnaît d'ailleurs la fécondité de cette étude des petits circuits élémentaires et promeut l'intégration de ces deux types d'approches. Nous allons cependant délaisser les travaux d'Uri sur les motifs dans la suite de ce chapitre : en effet, notre objectif n'est pas de discuter de la pertinence de la théorie des réseaux pour les réseaux biologiques en général ou de comparer les approches *bottom-up* et *top-down* dans le domaine de la biologie des réseaux. Notre objectif est spécifiquement d'étudier en quoi la médecine des réseaux peut constituer un cadre pour le développement d'une théorie génétique de la maladie et c'est principalement l'équipe de Barabási qui a discuté ce passage de la biologie des réseaux à la médecine des réseaux (Barabási, 2007 ; Barabási *et al*., 2011).

6.1.4. La médecine des réseaux, « une approche de la maladie humaine fondée sur les réseaux »

À l'ère post-génomique, la médecine des réseaux naît donc de la confluence entre la naissance de la biologie et de la médecine des systèmes, le concept de « gènes humains de la maladie » et le développement de la théorie des réseaux et de son application en biologie. Nous avons déjà évoqué le premier texte de 2007 de Barabási

où apparaît le terme de « médecine des réseaux ». Mais le texte fondateur est plutôt son article de 2011, intitulé « Network medicine, a network-based approach to human disease » et publié dans *Nature Genetics Reviews* (Barabási *et al.*, 2011). Remarquons d'abord que Barabási utilise le terme « d'approche de la maladie humaine fondée sur des réseaux » : il s'agit bien d'un programme de recherche, de l'utilisation d'une méthode spécifique (l'analyse des réseaux) pour comprendre différemment la maladie. En effet, la médecine des réseaux consiste principalement à considérer que les maladies ne sont pas des entités isolées qui peuvent être étudiées indépendamment les unes des autres, mais qu'il existe au contraire une forte interdépendance entre les différents phénomènes physiopathologiques. Dès lors, la meilleure méthode pour analyser les interactions qui existent entre les causes, les symptômes, l'évolution et le traitement des maladies est de considérer les différents réseaux impliqués dans la formation des processus pathologiques.

Pour cela, deux grands types de réseaux sont comparés : l'interactome et les réseaux des maladies (voir Figure 43). Au sens le plus strict du terme, on appelle interactome l'ensemble des interactions moléculaires qui existent au sein d'une cellule donnée à un moment donné. Il s'agit donc à la fois de l'ensemble des interactions gène-gène, gène-protéine, protéine-protéine, facteurs de transcription-protéine, etc. Mais au sens plus large, on parle d'interactome pour l'ensemble des interactions moléculaires qui existent au niveau d'un organisme dans des conditions données. La différence entre l'interactome d'une cellule et l'interactome d'un organisme est considérable puisqu'un organisme est constitué de différents types cellulaires (plus de deux sens chez l'homme) et qu'à l'échelle d'un organisme, il faut aussi considérer les interactions qui peuvent s'établir au sein et entre plusieurs niveaux d'organisation. Bien que plusieurs projets de cartographies des interactomes soient aujourd'hui en cours, aucun interactome n'est aujourd'hui complet ni pour une cellule, ni pour un organisme donné. L'interactome humain reste ainsi davantage un concept qu'une réalité. Les données sont partielles, difficiles à vérifier étant donné la variété des méthodes de collection et de stockage des données et concernent différents types de molécules et différents types d'interactions :

> « L'ensemble complet de ces interactions chez les humains est appelé l'interactome humain. La complexité d'un tel réseau est écrasante, puisque les humains ont à peu près 25 000 gènes codant des protéines et un nombre inconnu de protéines à cause du grand nombre de variants épissés et des modifications post-traductionnelles. Ainsi, on estime que le nombre de protéines qui sont impliquées dans l'interactome est un chiffre de l'ordre de 10^6. L'état actuel des bases de données est estimé à environ 50 000 protéines uniques qui participent à près de 200 000 interactions. Quand le score de qualité des mesures est pris en compte, les données des interactions protéines-protéines humaines tombent à environ 10 000 protéines participant à quelque 50 000 à 60 000 interactions estimées avec un intervalle de confiance élevé. Ceci montre que les données aujourd'hui disponibles sur l'interactome sont encore approximatives et incomplètes. De nombreux biais sont introduits par les techniques de collections et d'échantillonnages des données, ainsi qu'en moyennant la population d'une même espèce lorsqu'on utilise des modèles universels de génome et d'interactome. » (Janjic et Przulj, 2012, p. 525‑526)

Quant aux « réseaux des maladies », ils désignent l'ensemble des réseaux d'interactions moléculaires qui se produisent dans toutes les maladies. On y retrouve donc la même variété de réseaux que dans l'interactome (réseaux de gènes de la maladie, réseaux métaboliques de la maladie, etc.). La médecine des réseaux est donc une analyse de la maladie humaine qui repose soit sur la comparaison de l'interactome (réseau des interactions moléculaires non pathologiques) aux réseaux de la maladie, soit sur la comparaison et l'intégration des réseaux de la maladie entre eux (quels rapports entretiennent par exemple le réseau des gènes de la maladie avec le réseau des protéines de la maladie ?).

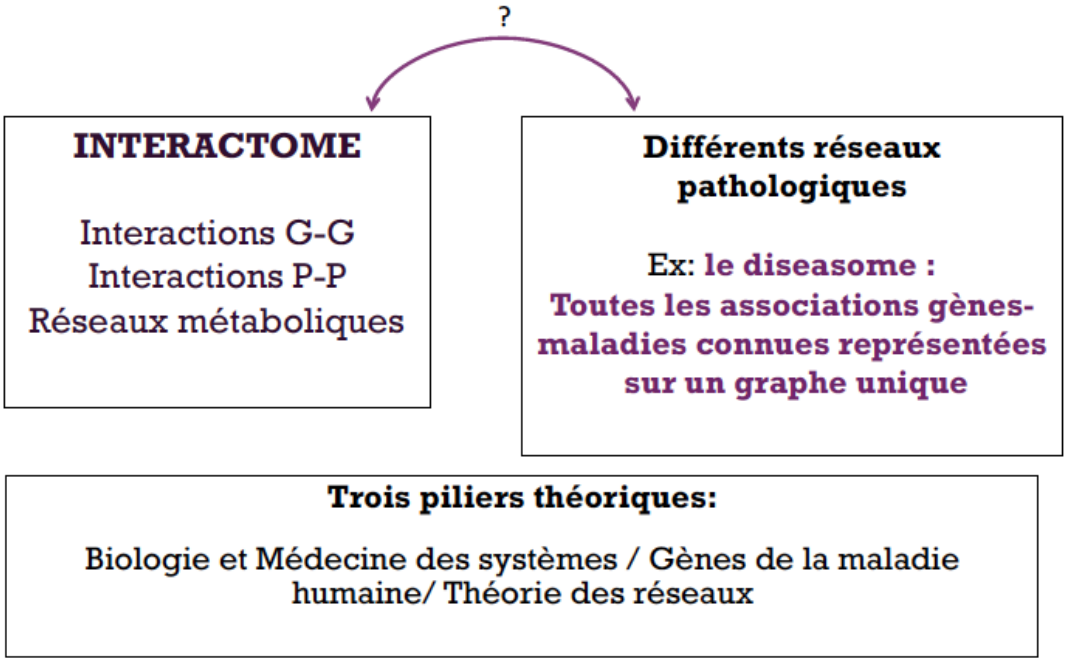

?
INTERACTOME
Interactions G-G
Interactions P-P
Réseaux métaboliques
Différents réseaux pathologiques
Ex: le diseasome :
Toutes les associations gènes-maladies connues représentées sur un graphe unique
Trois piliers théoriques:
Biologie et Médecine des systèmes / Gènes de la maladie humaine/ Théorie des réseaux

Figure 43 : Représentation des fondements conceptuels de la médecine des réseaux. La médecine des réseaux repose sur trois piliers théoriques : la biologie et la médecine des systèmes, le concept de gènes de la maladie humaine et la théorie des réseaux. La médecine des réseaux consiste à comparer l'ensemble des interactions moléculaires à l'échelle d'un organisme donné ou dans une maladie donné à différents réseaux pathologiques.

Notre objectif est de déterminer à quel point la médecine des réseaux est un programme de recherche dans lequel peut se développer une théorie génétique de la maladie. Par conséquent, dans la suite de ce chapitre, nous allons principalement nous concentrer sur un seul type de réseau : le diseasome ou réseau des gènes humains de la maladie.

6.2. Construire et analyser le diseasome

À présent que nous avons exposé le cadre historique et intellectuel dans lequel est née la médecine des réseaux, il nous faut exposer en détail la construction et l'analyse d'une des réalisations principales de la médecine des réseaux – le diseasome, tel qu'il a été présenté dans un article fondateur de 2007 intitulé « The human disease network » (Goh *et al.*, 2007).

Le diseasome ou « réseau des gènes de la maladie humaine » est une représentation à la fois très simple et très ambitieuse, sous forme de réseaux, de toutes les interactions connues entre des maladies et les gènes impliqués dans ces maladies. Il est important de remarquer que le but explicite et final est d'intégrer le diseasome à l'interactome. Deux autres objectifs sont poursuivis. Le premier est d'obtenir un outil de visualisation qui permette d'identifier des interactions passées inaperçues jusqu'ici entre deux maladies afin de diriger de façon plus efficace la recherche de gènes candidats et de comprendre la façon dont ces gènes interagissent. Le second est d'explorer les propriétés spécifiques des gènes de la maladie humaine et de les comparer avec les autres gènes humains.

Notons que parmi les auteurs de cet article, on retrouve plusieurs auteurs que nous avons déjà cités dans la première section de ce chapitre : Barabási, bien sûr, mais aussi David Valle et Barton Childs qui avaient participé aux premiers articles sur le concept de gènes humains de la maladie, Marc Vidal dont l'équipe a publié des travaux pionniers sur l'interactome dès le début des années 2000 (Walhout et al., 2000). Dans un article de 2012, qui introduit les principes de l'analyse du diseasome, Kwang Il-Goh (premier auteur de l'article de 2007, professeur associé au département de physique de l'Université de Corée, spécialisé dans l'analyse des réseaux complexes) revient sur les motivations principales de la construction du diseasome :

> « A cause de la complexité intrinsèque des associations génotype-phénotype, la cause et les effets d'une maladie deviennent de plus en plus ambigus et il est difficile d'élucider les mécanismes sous-jacents. Ainsi, organiser les associations individuelles gènes-maladies devient de plus en plus compliqué et la nécessité d'une vision globale des relations entre les maladies et leurs composants génétiques est devenue inévitable. Sur cette nouvelle vague de la génétique des maladies, la théorie des réseaux complexes, une branche des mathématiques et de la physique théorique, joue un rôle instrumental en fournissant de nouvelles pistes conceptuelles, ainsi qu'en offrant une méthodologie visuelle et computationnelle. » (Goh et Choi, 2012, p. 535)

On retrouve dans cette citation tous les ingrédients qui sont à l'origine de la médecine des réseaux : étudier les maladies non pas isolément mais comme un

ensemble, faire sens de l'afflux et de la disparité des données post-génomiques, utiliser la théorie des réseaux comme un instrument de visualisation et d'analyse de ces données.

Nous allons d'abord détailler la façon dont les auteurs ont construit le diseasome (section 6.2.1.), avant de présenter l'analyse globale (section 6.2.2) et locale (section 6.2.3.) du diseasome. Enfin, nous passerons à la comparaison du diseasome avec l'interactome (section 6.2.4.).

6.2.1. Construire le diseasome

Pour construire le diseasome, les auteurs ont procédé en deux étapes (voir Figure 44) : (1) La première étape consiste à construire un graphique bipartite avec deux ensembles disjoints de nœuds, représentant d'un côté, sous forme de cercles, l'ensemble des maladies humaines connues *(disease phenome)* et de l'autre, sous forme de rectangles, l'ensemble des gènes dont l'implication dans une maladie est connue (*disease genes*). Sur ce graphique bipartite (au centre de la Figure 44), un gène est connecté à une maladie si des mutations dans le gène sont impliquées dans la maladie. Pour construire ce graphique, les auteurs ont utilisé *Online Mendelian Inheritance on Man*, la base de données que nous avons déjà évoquée en détail dans le chapitre 1 et qui, en décembre 2005, contenait 1 284 maladies et 1 777 gènes. (2) À partir de ce graphique bipartite, on peut construire deux projections : le réseau des maladies humaines où les nœuds sont des maladies et où deux maladies sont connectées si leur physiopathologie implique un gène commun (à gauche sur la Figure 44) et le réseau des gènes de la maladie où les nœuds sont des gènes et où deux gènes sont connectés si ils sont impliqués dans la même maladie (à droite sur la Figure 44). La taille des cercles est proportionnelle au nombre de gènes impliqués dans la maladie correspondante. L'épaisseur du lien (ou arête) est proportionnelle au nombre de maladies dans lesquelles deux gènes sont impliqués. Enfin, une couleur est attribuée en fonction de la classe de maladie à laquelle appartient la maladie concernée. Vingt-deux classes de maladie ont ainsi été identifiées (la plupart du temps selon leur localisation anatomoclinique) et toutes les maladies ont été classées manuellement dans une de ces classes. On notera l'existence d'une classe « multiple » quand une maladie concerne plusieurs localisations anatomo-cliniques, l'existence d'une

catégorie « inclassable » et l'existence de catégories étiologiques plutôt qu'anatomo-cliniques (classe des maladies immunologiques, par exemple).

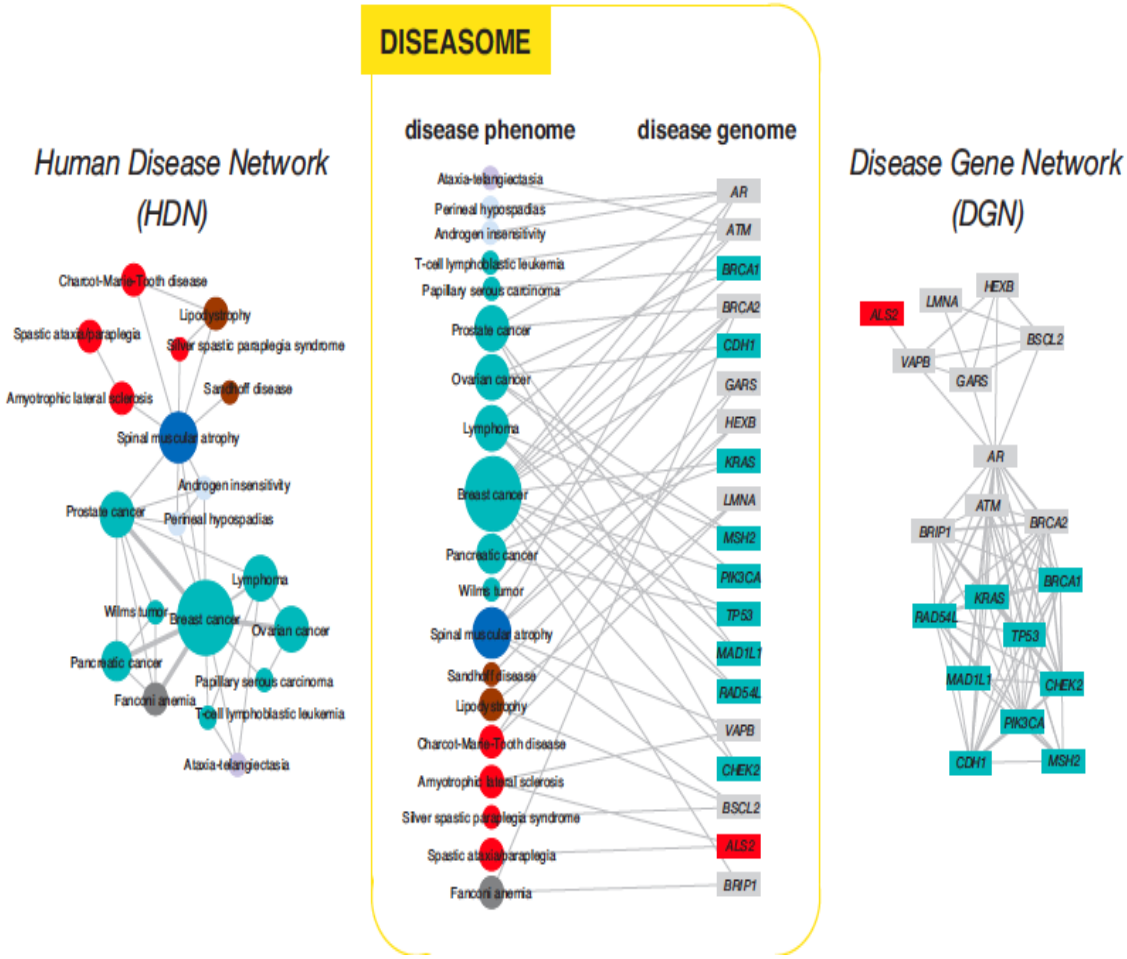


Figure 44 : Construction du diseasome (Goh et al., 2007, p. 8686). Le diseasome est construit à partir d'un graphe bipartite qui répertorie les associations entre gènes et maladies. A gauche, le réseau des maladies humaines est un réseau dont les nœuds sont des maladies. Deux maladies sont liées par une arête si elles partagent le même gène dans leur physiopathologie. A droite, le réseau des gènes humains de la maladie est un réseau dont les nœuds sont des gènes. Deux gènes sont liés par une arête s'ils sont impliqués dans la même maladie.

La Figure 44 ne représente qu'une petite partie de l'échantillon du diseasome, la représentation graphique complète est ci-dessous (Figure 45).

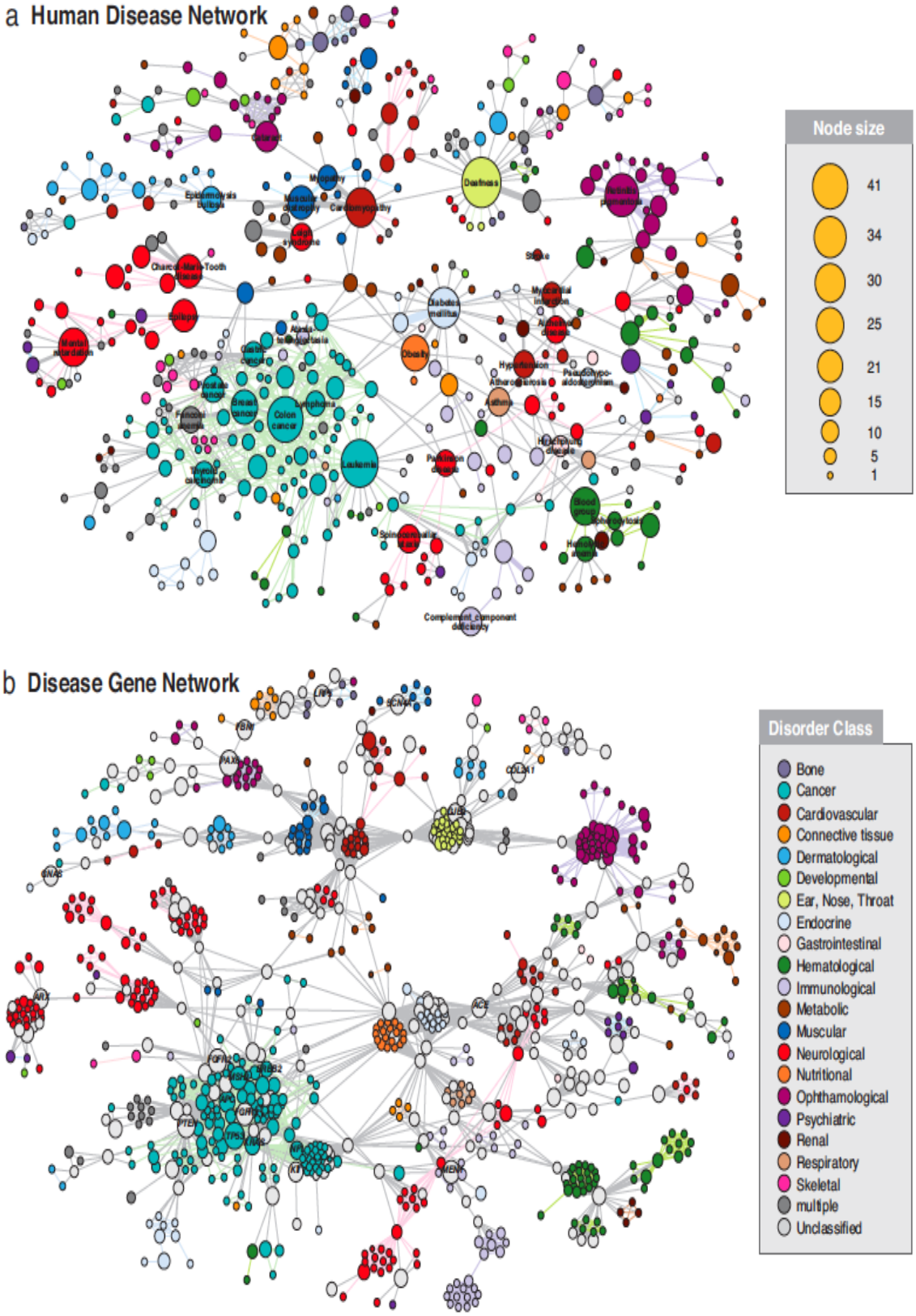


Figure 45 : Diseasome - A. Le réseau des maladies humaines. B. Le réseau des gènes de la maladie humaine (Goh et al., 2007, p. 8687). La taille des nœuds est proportionnelle au degré de connectivité. Le code couleur permet de distinguer différentes classes de maladie.

Sur la Figure 45, on remarquera que dans le réseau humain de la maladie (Figure 45A), lorsque les arêtes connectent des maladies apparentant à une même classe de maladie, le lien apparaît dans une couleur correspondant à la classe de maladie en question. En revanche lorsque les liens connectent des maladies appartenant à différentes classes de maladie, les liens sont représentés en gris. Les noms des maladies sont indiqués lorsqu'une maladie est associée à plus de dix gènes. Dans le réseau des gènes humains de la maladie (Figure 45B), les nœuds sont en gris clair, si les gènes correspondants sont associés avec plus d'une classe de maladie. Les gènes associés à plus de cinq maladies sont nommés. Seuls les nœuds ayant au moins un lien sont montrés. Précisons également que le diseasome n'existe pas seulement sous forme de texte : les auteurs ont par ailleurs développé une version interactive en ligne du diseasome[56].

On peut d'ores et déjà identifier un certain nombre de limites dans la construction du diseasome. D'abord, la reconstruction du réseau de la maladie humaine comme du réseau des gènes de la maladie humaine sont dépendants des données utilisées. En l'occurrence, les auteurs ont fait le choix d'utiliser la base de données *Online Mendelian Inheritance On Man*, dont on a plusieurs fois souligné qu'elle ne s'était ouverte que très récemment aux maladies polyfactorielles. Par ailleurs, il est bien évident que le diseasome est par nature incomplet et qu'il a vocation à s'agrandir : ne sont indiquées que les associations gènes-maladies qui sont connues en 2005. Dès lors, on peut légitimement se demander à quel point le diseasome est un bon modèle des relations entre les maladies et leurs composants génétiques et à quel point les analyses qui en sont tirées sont extrapolables aux très nombreuses associations qu'il nous reste à découvrir. Dans la discussion de leur article de 2007, les auteurs répondent en partie à ces objections par une simulation expérimentale : ils ont choisi de prendre en compte non pas seulement les associations gènes-maladies où des gènes avec des mutations spécifiques bien identifiées ont été associés à une maladie, mais aussi les associations où le phénotype n'a pas été spécifiquement associé à une mutation précise (mais qui sont aussi répertoriées dans

[56] Consultable à l'adresse suivante : http://diseasome.eu/index.html.

OMIM, comme nous le décrivions dans le chapitre 1). Malgré l'insertion de ces nouvelles données, la structure générale du réseau ne change pas :

> « Pour tester si les conclusions obtenues dans ce travail sont robustes malgré le caractère incomplet d'OMIM, nous avons étendu notre étude pour inclure non seulement les gènes avec des mutations identifiées liés à un phénotype spécifique, mais aussi ceux qui satisfont un critère moins exigeant, lorsque le phénotype n'a pas été identifié à un locus spécifique. Cette expansion a augmenté le nombre de gènes associés à des maladies de 1 777 à 2 765, mais a aussi introduit du bruit dans les données, puisque le lien entre de nombreux nouveaux gènes introduits et les maladies correspondantes est moins rigoureux. Cependant, la structure générale de ce diseasome étendu demeure largement stable et aucune des tendances découvertes […] n'est affectée par cette expansion possible de la base de données OMIM. Ainsi, bien que les cartes montrées dans la Figure 2 [correspondant à notre Figure 45] subiront inévitablement des changements locaux lors de la découverte de nouveaux gènes de la maladie, cela ne changera pas l'organisation générale et la distribution de réseau des maladies humaines de façon significative […] . » (Goh et al., 2007, p. 8690)

A cette confirmation *in silico* de la fiabilité du diseasome tel qu'il a été construit s'ajoute une propriété théorique intéressante de la science des réseaux : les propriétés des réseaux invariants d'échelle sont des « *overdetermined properties* », il n'est théoriquement pas nécessaire de connaître la totalité des nœuds du réseau pour en identifier le fonctionnement. Enfin, depuis la publication de l'article, une autre équipe a proposé de recréer une version étendue du diseasome (Zhang *et al.*, 2011) ne prenant pas seulement en compte les interactions entre les maladies et les gènes, mais aussi entre les maladies et les protéines. Cette version s'appuie sur une base de données différente, la base de données GAD *Genetic Association Database*[57] (Becker *et al., 2004).* La structure générale et les propriétés topologiques de cette version étendue du diseasome qui s'appuie sur les données de GAD sont parfaitement compatibles et très proches de la version initiale du diseasome s'appuyant sur les

[57] http://geneticassociationdb.nih.gov/

données OMIM. C'est donc une forme de confirmation expérimentale de la crédibilité du diseasome.

Il s'agit maintenant de présenter les trois grandes analyses que proposent les auteurs du diseasome : la première analyse (section 6.2.2.) est globale et topologique, la seconde analyse est locale et fonctionnelle (section 6.2.3.), la troisième consiste à comparer le diseasome à l'interactome (section 6.2.4.). Chaque type d'analyse utilise des méthodes différentes, faisant parfois appel à des données extérieures (c'est-à-dire des données qui ne dépendent pas de la structure du réseau et qui viennent parfois d'autres bases de données qu'OMIM). Il nous semble important d'insister sur ce point, à la fois pour comprendre la diversité des méthodes mises en œuvre dans la médecine des réseaux et pour en souligner les éventuelles limites.

6.2.2. Analyser globalement le diseasome comme un réseau invariant d'échelle

La première analyse du diseasome est globale et topologique : il s'agit de qualifier le comportement général et la topologie des deux réseaux, en utilisant les outils de la théorie des réseaux. Les auteurs remarquent que dans le réseau des maladies humaines, comme dans le réseau des gènes humains de la maladie, les nœuds (respectivement, les maladies et les gènes) sont fortement interconnectés (très peu de nœuds ne sont absolument pas reliés au réseau général) et que la distribution des degrés dans le réseau des maladies humaines comme dans le réseau des gènes humains de la maladie se comportent comme des réseaux invariants d'échelles avec des maladies ou des gènes jouant le rôle de *hubs* hyper connectés, tandis que la plupart des autres nœuds ont un degré faible.

Ainsi, dans le réseau des maladies humaines, sur 1 284 maladies, 867 sont reliées à une autre maladie et 516 (soit 40% des maladies représentées) forment un seul groupe (*« cluster »)* bien identifiable. Certaines maladies, comme le cancer, jouent donc le rôle de connecteurs (*hubs*). Ainsi, le cancer du colon est lié à 50 autres maladies et le cancer du sein à 30 autres maladies. Dans le réseau des gènes humains de la maladie, sur 1 777 gènes de la maladie, 1 377 gènes sont connectés à d'autres gènes de la maladie et 903 gènes appartiennent à un composé géant. Alors que le nombre de gènes impliqués dans de nombreuses maladies décroit rapidement (la plupart des gènes sont connectés à un petit nombre de maladies), certains gènes de la

maladie (comme TP53 par exemple) sont impliqués dans plus de dix maladies, représentant des *hubs* majeurs dans le réseau des gènes humains de la maladie. Inversement, la plupart des maladies sont associées à quelques gènes, mais quelques maladies sont associées à plus de trente gènes. Ainsi, la surdité est associée à 41 gènes, la leucémie à 37 et le cancer du colon à 34.

La distribution des maladies dans le réseau des maladies humaines correspond relativement bien à la distribution des classes de maladie, mais il y a des différences notables en fonction des classes de maladie. Ainsi, si les maladies cancéreuses sont fortement interconnectées, du fait de l'implication commune de nombreux gènes dans plusieurs sous-types de cancers, d'autres classes de maladie comme les maladies métaboliques sont sous-représentées dans le groupe géant mais sur-représentées dans les composants plus petits (Figure 46).

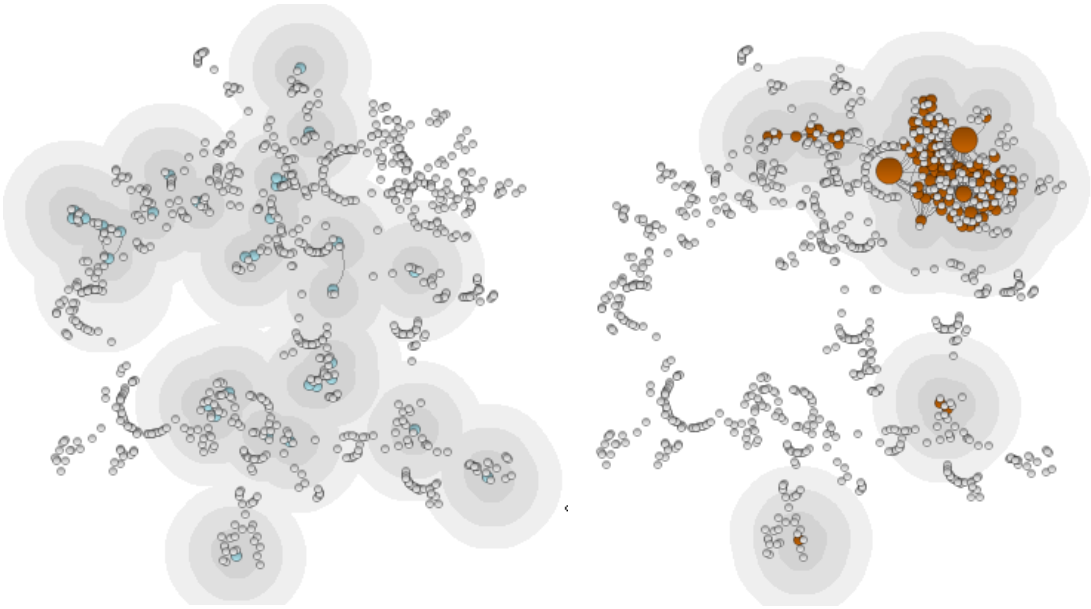


Figure 46 : Différence de connectivité selon les classes de maladie. Dans ces deux représentations tronquées du diseasome du réseau des maladies humaines, on a choisi de représenter les liens maladie-maladie d'une même classe de maladie. A gauche, seuls les liens entre maladies métaboliques (nœuds bleus) ont été représentés. A droite, seuls les différents sous-types de cancers (nœuds de couleur marron) ont été représentés. On remarque le contraste entre les maladies métaboliques qui sont très dispersées et peu connectées entre elles et les cancers qui forment un grand groupe. **Source : http://diseasome.eu/**

L'hypothèse retenue pour expliquer cette différence de connectivité entre classes de maladie est la différence d'hétérogénéité génétique (le fait que des

mutations sur plusieurs gènes différents puissent causer des phénotypes différents). Ainsi, le cancer et les maladies neurologiques qui ont une forte hétérogénéité allélique sont extrêmement interconnectés, tandis que les maladies métaboliques, orthopédiques, et les désordres systémiques (à multiples localisations anatomocliniques) qui montrent une faible hétérogénéité sont moins connectés entre eux.

L'hypothèse des auteurs est que, si chaque maladie humaine avait une origine génétique distincte et unique, le réseau des maladies humaines serait soit complètement déconnecté et composé de nombreux nœuds isolés correspondant à chaque maladie, soit composé d'un ensemble de petits groupes de désordres similaires. Le fait que la distribution des deux réseaux ne corresponde à aucun de ces deux schémas et qu'il ait été mesuré que leur distribution s'écartait significativement d'une distribution aléatoire suggère au contraire une origine génétique commune pour de nombreuses maladies et l'existence d'un regroupement fonctionnel et modulaire des maladies ou des gènes au sein du réseau.

6.2.3. Analyser localement le diseasome : l'hypothèse locale

La deuxième analyse est globale et fonctionnelle : il s'agit justement de tester cette hypothèse d'un regroupement fonctionnel des gènes de la maladie humaine et d'analyser le comportement et les propriétés de gènes qui sont impliqués dans une même maladie. Plus précisément, la première analyse du diseasome semble indiquer que les gènes et les maladies ne sont pas associés par hasard. L'hypothèse à tester est donc que les gènes qui sont impliqués dans une même maladie font partie d'un même module fonctionnel de la maladie donnée et vice-versa, que deux maladies qui partagent le même gène ont en commun une partie de leur physiopathologie. Pour tester cette hypothèse locale, il faut donc caractériser si les gènes impliqués dans une même maladie ont des produits (comme les protéines) qui interagissent ensemble, s'ils sont exprimés dans des tissus similaires, s'ils sont exprimés en même temps, s'ils ont des fonctions moléculaires proches.

Pour ce faire, il faut introduire un certain nombre de connaissances biologiques sur les gènes et les maladies impliqués dans le diseasome. C'est ainsi que les auteurs ont utilisé les informations venant d'OMIM sur chaque gène et chaque maladie du

réseau, mais aussi : a) un réseau d'interactions protéine-protéine, contenant 22 053 interactions (excluant les auto-corrélations et les redondances) entre 7 533 gènes et qui a été obtenu en combinant les résultats de deux expériences de technique systématique de double hybride de levure avec une liste d'interactions protéine-protéine obtenues par une revue manuelle de la littérature, b) les annotations GO pour chaque gène. Rappelons que le projet *Gene Ontology* est un projet de description standardisée des produits des gènes. L'objectif est de caractériser les produits d'un gène en décrivant à la fois les composants cellulaires (intra ou extracellulaire) avec lesquels ils interagissent, les fonctions moléculaires qu'ils réalisent (par exemple une réaction de phosphorylation pour une protéine) et les processus biologiques dans lesquels ils sont impliqués. c) des données sur l'expression des gènes du réseau[58] (à quel moment, en quelle quantité et dans quels tissus ?), obtenues à partir de l'analyse de puces à ADN ou à ARN qui sont répertoriées dans la base de données Entrez gene ID qui est reliée à OMIM .

En comparant ces données biologiques aux données du diseasome, les auteurs concluent que les gènes impliqués dans une même maladie ont tendance à : (a) interagir via des interactions protéines – protéines, (b) être exprimés dans des tissus spécifiques, (c) être fortement coexprimés, c'est à dire être exprimés en même temps (d) être exprimés de façon synchronisée comme un groupe synchrone, (e) partager les mêmes termes GO. Cette hypothèse locale permet d'introduire la notion de module fonctionnel commun à une même maladie :

> « Les réseaux cellulaires sont modulaires et sont constitués de groupes de protéines hautement interconnectées responsables de fonctions cellulaires spécifique. Ainsi, une maladie représente alors la perturbation ou la rupture d'un module fonctionnel spécifique causée par la variation d'un ou de plusieurs composants produisant des anomalies reconnaissables du développement et/ou de la physiologie. Ce modèle offre une explication fondée sur les réseaux pour l'émergence des maladies complexes ou polygéniques : un phénotype est souvent corrélé à l'incapacité d'un module fonctionnel particulier à réaliser ses fonctions fondamentales. Pour les modules étendus, de nombreuses combinaisons différentes de gènes

[58] Ces données n'étaient disponibles que pour 1 357 gènes, soit 70% des gènes du diseasome.

> perturbés peuvent empêcher le module de fonctionner, en conséquence de quoi les mutations de gènes différents vont sembler conduire au même phénotype. Cette corrélation entre la maladie et les modules fonctionnels peut aussi enrichir notre compréhension des réseaux cellulaires en nous aidant à identifier quels gènes sont impliqués dans une même fonction cellulaire ou dans un même module. » (Goh et al., 2007, p. 8688)

C'est là une des affirmations les plus fortes de la médecine des réseaux sur la génétique des maladies et une conséquence logique de l'hypothèse de l'origine génétique commune des maladies humains : si des maladies partagent des mêmes gènes ou que plusieurs gènes sont associés à la physiopathologie d'une même maladie, c'est qu'ils appartiennent à un module fonctionnel, c'est-à-dire à un groupe d'éléments moléculaires pouvant comprendre des facteurs de transcription, des gènes, des protéines, etc., qui interagissent d'une certaine manière et réalisent une fonction cellulaire ou moléculaire donnée. Nous commenterons un peu plus loin et plus en détail cette notion de module fonctionnel.

6.2.4. Analyser le diseasome au sein de l'interactome : propriétés des gènes humains de la maladie

La troisième analyse consiste à comparer le diseasome à l'interactome : en comparant les propriétés des gènes impliqués dans les maladies humaines à ceux dont l'implication n'est pas suspectée dans des processus physiopathologiques, les auteurs espèrent acquérir un certain nombre de propriétés caractéristiques des gènes humains de la maladie. Rappelons à ce propos que l'interactome n'est aujourd'hui qu'incomplet et partiel (en l'occurrence, les auteurs ont approximé l'interactome par le réseau étendu d'interactions protéine-protéine dont nous avons décrit les conditions d'obtention dans la section précédente).

Une des propriétés les plus étudiées est la question de la centralité-essentialité des gènes. En effet, lorsque les réseaux cellulaires de la levure ont été analysés par l'équipe de Barabási en 2001, une des conclusions était que les protéines les plus hautement connectées et qui jouaient le rôle de *hubs* dans le réseau des interactions protéine-protéine de la levure avaient tendance à encoder des gènes dits essentiels. Le concept de « gène essentiel » est un concept intrinsèquement lié aux expériences de

mutants *knock-out* (c'est-à-dire de modèles animaux où l'on inactive de façon systématique un seul gène à la fois, afin d'observer l'effet de ce gène sur le phénotype), puisque sont considérés comme essentiels les gènes qui sont nécessaires à la survie d'un organisme, c'est-à-dire les gènes qui dans les expériences de *knock-out* entrainent la mort du phénotype mutant. Une telle définition n'a rien d'évident, étant donné que l'essentialité d'un gène peut dépendre des conditions environnementales dans lesquelles l'organisme se développe et en l'occurrence des conditions expérimentales dans lesquelles est étudié l'effet du *knock-out*. Un concept relativement proche est celui de « gène de ménage » (*housekeeping gene*) : un gène de ménage est un gène qui est impliqué directement ou par ses produits dans des fonctions critiques de la cellule (cytosquelette, cycle du glucose, etc.) et dont le niveau et le lieu (type de tissu) d'expression sont constants dans le temps. En l'occurrence, pour prédire l'essentialité d'un gène humain à partir du réseau étendu d'interactions protéine-protéine, les auteurs ont utilisé les informations disponibles pour les gènes orthologues de la souris, à partir de la base de données Mouse Genome Informatics[59]. Plus spécifiquement, un gène humain a été défini comme essentiel si le *knock-out* de son orthologue murin entrainait la mort du phénotype mutant correspondant, que ce soit à l'état embryonnaire, à l'état prénatal ou pendant la période postnatale immédiate (tous les autres phénotypes ont été considérés comme non-létaux).

En se fondant sur l'analyse du réseau des interactions protéine-protéine de la levure, les auteurs s'attendaient donc à ce que les gènes humains de la maladie soient des gènes essentiels selon leur définition et constituent des *hubs* au sein de l'interactome. Une première analyse du réseau étendu d'interactions protéine-protéine (sans utiliser les orthologues murins correspondants) montre que les protéines issues des gènes de la maladie ont tendance à avoir une connectivité (nombre de degrés) plus importante (c'est-à-dire qu'elles ont 32% d'interactions supplémentaires) par rapport aux protéines dont les gènes ne sont pas impliqués dans la maladie humaine et que les protéines avec un degré élevé (qui ont de nombreuses interactions) ont davantage tendance à être encodées par des gènes associés à des maladies que les protéines avec un degré moins important (qui ont un petit nombre

[59] http://www.informatics.jax.org

d'interactions). Si on s'arrête à une corrélation entre la capacité d'être un *hub* et le caractère essentiel d'une protéine, cette première analyse semble confirmer l'hypothèse selon laquelle les gènes de la maladie sont des *hubs* essentiels au sein de l'interactome.

Cependant, une analyse plus poussée montre des situations très variées entre différents types de gènes de la maladie. En utilisant les orthologues murins pout déterminer le niveau d'essentialité des gènes du réseau étendu d'interactions protéine-protéine, les auteurs ont identifié sur les 7 533 gènes du réseau étendu d'interactions protéine-protéine 1 267 gènes essentiels et non associés à la maladie et ont classé les 1 777 gènes de la maladie humaine en deux catégories : 398 gènes de la maladie humaine essentiels et 1 379 gènes de la maladie humaine non essentiels (Figure 47).

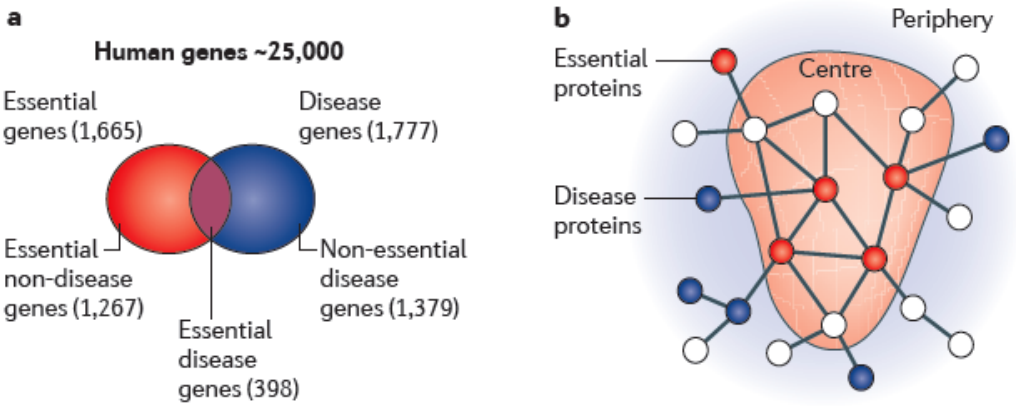


Figure 47 : Les gènes de la maladie au sein de l'interactome (Barabási *et al.*, 2011, p.59*)*. Sur les 20 à 25 000 gènes humains, 2 418 sont associés à des maladies spécifiques. La figure A montre un chevauchement entre les 1 777 gènes associés à des maladies (connus en 2007) et les 1 665 gènes qui sont considérés comme essentiels *in utero*, c'est-à-dire dont l'absence est associée à la perte de l'embryon. La figure B est une représentation schématique des différences entre les gènes essentiels et les gènes non-essentiels. Les gènes non essentiels (en bleu) ont tendance à être à la périphérie du réseau, alors que les gènes essentiels *in utero* (en rouge) occupent le centre fonctionnel de l'interactome, encodant des *hubs* et s'exprimant dans de nombreux tissus.

Ils ont ensuite remarqué que les protéines essentielles (qu'elles soient ou non associées à la maladie) avaient une très forte tendance à former des *hubs* et ils se sont demandé si les résultats de leur première analyse n'avaient pas été faussés par l'inclusion des gènes essentiels de la maladie (qui ne représentent pourtant qu'une petite partie de l'ensemble des gènes de la maladie connus). En mesurant le degré de

corrélation entre les protéines non essentielles mais associées à la maladie et la tendance à former des *hubs*, la corrélation précédente avait complètement disparu.

Ces résultats, associés à d'autres analyses sur le niveau et le lieu d'expression de ces gènes non essentiels de la maladie ont permis de conclure, contre toute attente, que la grande majorité des gènes non essentiels de la maladie occupe une place périphérique dans le réseau cellulaire et ne sont exprimés que dans un petit nombre de tissus. Par contraste, les gènes essentiels ont tendance à encoder des hubs, montrent une expression très synchronisée avec le reste des gènes et sont exprimés dans la majorité des tissus. Comment comprendre cette non-essentialité de la majorité des gènes de la maladie humaine ? Les auteurs font ici une hypothèse évolutionnaire :

> « Ce caractère périphérique inattendu de la majorité des gènes de la maladie peut être expliqué en utilisant un argument évolutionnaire. Les mutations dans des gènes topologiquement centraux et largement exprimés ont plus de chances de constituer un obstacle sévère au développement et à des fonctions physiologiques normaux, entrainant la mort *in utero* ou au début de la vie extra-utérine et éventuellement la disparation de cette mutation dans la population. Seules les mutations compatibles avec la survie jusqu'à l'âge de la reproduction ont une chance de se maintenir dans la population. C'est pourquoi les mutations associées aux maladies dans des régions périphériques de la cellule d'un point de vue topologique et d'un point de vue fonctionnel, donnent de meilleures chances de viabilité. » (Goh *et al.,* 2007, p. 8690)

Dans cette logique, on comprend pourquoi la majorité des gènes de la maladie sont non essentiels et situés à la périphérie de l'interactome : le petit groupe de gènes de la maladie qui sont au contraire essentiels est probablement responsable de la majorité des anomalies prénatales fatales et des décès *in utero*, empêchant le maintien de la mutation dans la population.

Nous avons présenté en détail cet article à la fois parce qu'il a joué un rôle fondateur dans la mise en place des concepts de la médecine des réseaux et parce qu'il illustre bien les combinaisons de méthodes (science des réseaux, comparaison de réseaux, etc.) et de données extrêmement variées (utilisation de modèles animaux, combinaisons de données venant de bases de données ou d'expériences différentes, utilisation des ontologies biomédicales) qui fondent à la fois l'élégance et les limites de

l'analyse de la génétique des maladies dans la médecine des réseaux. Nous allons maintenant approfondir notre description des concepts fondamentaux (définition modulaire de la maladie) de la médecine des réseaux, en analysant à quel point elle peut constituer un cadre au développement d'une théorie génétique générale de la maladie.

6.3. En quoi la médecine des réseaux est-elle un cadre pour une théorie génétique générale de la maladie ?

Nous parlons de la médecine des réseaux comme d'un « cadre » pour le développement d'une théorie génétique générale de la maladie pour deux raisons. (1) D'une part, il faut garder à l'esprit que la médecine des réseaux ne se réduit pas à l'analyse des réseaux gènes-maladies, mais comprend aussi l'analyse des réseaux de métabolites, ou encore des réseaux de protéines qui sont impliqués dans les maladies. C'est cependant sur l'analyse des réseaux de gènes de la maladie que nous nous focaliserons dans la suite, car c'est bien elle qui est fondamentale pour notre propos. (2) D'autre part, contrairement à Casanova et Abel pour la théorie génétique des maladies infectieuses, les partisans de la médecine des réseaux n'emploient jamais explicitement le terme de « théorie génétique » pour qualifier leurs travaux : c'est bien nous qui reconstruisons leurs propos comme une théorie génétique générale de la maladie. Pour le dire encore différemment, nous avons insisté dans notre chapitre 4 sur la distinction entre deux types de théories générales médicales : une théorie générale des maladies qui est un regroupement hétérogène de théories régionales, chaque classe de maladie ayant sa propre théorie régionale et une théorie générale de la maladie qui s'appuie sur une définition globale de la maladie. Ce que nous voulons soutenir ici c'est que la médecine des réseaux, en s'appuyant sur une définition de la maladie en termes de module fonctionnel, est une théorie générale de la maladie. À l'intérieur de cette théorie générale, les travaux les plus approfondis concernent les réseaux génétiques de la maladie humaine et nous soutenons que ce cadre constitue un terrain fécond pour le développement d'une théorie génétique partielle de la maladie en général.

Dans la section 6.3.1., nous démontrerons en quoi la médecine des réseaux est une théorie générale de la maladie et en quoi cette théorie est une explication interactionniste co-constructionniste de la maladie, au sein de laquelle on peut trouver une théorie génétique partielle de la maladie. Dans la section 6.3.2., nous montrerons plus précisément que cette théorie génétique de la maladie touche les quatre cibles explicatives : causes, symptômes, évolution, traitement. Dans la section 6.3.3., nous discuterons la nature des explications que mobilise cette théorie génétique. Enfin, dans la section 6.3.4., nous reviendrons sur les trois critères supplémentaires que nous avons identifiés comme des indicateurs de précision d'une théorie : à quel point la théorie génétique de la maladie développée dans le cadre de la médecine des réseaux permet-elle d'expliquer la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité des maladies ?

6.3.1. La médecine des réseaux, une théorie générale interactionniste et co-constructionniste qui explique les causes, les symptômes, l'évolution et le traitement de la maladie

La médecine des réseaux est une théorie générale de la maladie pour plusieurs raisons. (1) Elle s'appuie sur une définition générale de la maladie applicable à toute maladie – ce faisant elle se rapproche de la structure des théories antiques dans lesquelles toute maladie-type peut être instanciée à partir d'une théorie générale. (2) D'ailleurs, comme chaque module de la maladie est spécifique d'une maladie donnée, une maladie est définie et identifiée par son module. (3) Mais si chaque module est spécifique d'une maladie, toute perturbation d'un module n'entraîne pas la maladie et tous les individus ont des modules différents.

6.3.1.1. Chaque maladie peut être définie par son module

Ce qui distingue la médecine des réseaux d'une théorie générale des maladies et qui en fait une théorie générale de la maladie, c'est qu'elle repose sur une définition modulaire de la maladie que nous avons précédemment citée et qui peut s'appliquer à toute maladie, que l'on parle d'une maladie-type ou d'une maladie instance :

> « Ainsi, une maladie représente alors la perturbation ou la rupture d'un module fonctionnel spécifique causée par la variation d'un ou de plusieurs composants produisant des anomalies reconnaissables du développement et/ou de la physiologie. » (Goh *et al.,* 2007, p. 8688)

Dans la revue de littérature de 2011, Barabási apporte une précision fondamentale sur le concept de module (Barabási *et al., 2011)*. Il distingue ainsi trois types de modules qui sont souvent implicitement confondus dans la littérature mais qu'il est nécessaire de distinguer en principe et en pratique : le module topologique, le module fonctionnel et le module de la maladie (voir Figure 48). Les modules topologiques (à gauche) sont définis comme des zones de forte densité locale au sein de l'interactome. Ils peuvent être identifiés par des algorithmes de réseaux qui sont aveugles à la fonction des gènes. Les modules fonctionnels (au centre) sont définis comme des réseaux voisins, dans lesquels il y a une agrégation statistiquement significative de nœuds (en gris) contribuant à une même fonction. Les modules de la maladie sont des groupes de nœuds qui, lorsqu'ils sont perturbés (mutations, délétions, variations du nombre de copies, changement d'expressivité), peuvent être liés à un phénotype pathologique particulier (représentés en rouge).

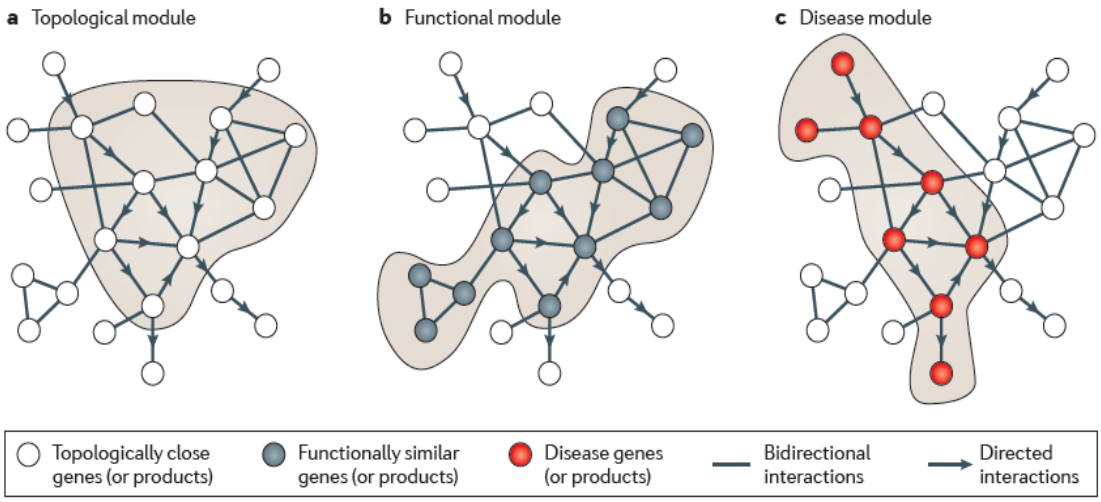


Figure 48 : Représentation schématique de trois concepts de modules de la maladie : le module topologique, le module fonctionnel et le module de la maladie (Barabási *et al.,* 2011, p. 61). Les trois modules ne se superposent pas strictement.

Comme on le voit sur la Figure 48, les module topologique, module fonctionnel et module de la maladie se chevauchent mais ne sont pas strictement identiques : quelles sont donc les relations entre ces trois concepts de modules ?

> « Dans la littérature biologique, il y a une hypothèse tacite selon laquelle les trois concepts sont reliés : les composants cellulaires qui forment un module topologique ont des fonctions reliées, c'est pourquoi ils correspondent aussi au module fonctionnel et une maladie résulte de la rupture d'un module fonctionnel particulier, ce qui sous-entend qu'un module fonctionnel est aussi un module de la maladie. Cependant, il faut garder à l'esprit certaines caractéristiques uniques des modules de la maladie. Tout d'abord, un module de la maladie peut ne peut pas être identique, mais va très probablement chevaucher, les modules topologiques et/ou les modules fonctionnels. Deuxièmement, un module de la maladie est défini en fonction d'une maladie particulière et, par conséquent, chaque maladie a son propre module unique. Enfin, un gène, une protéine ou un métabolite peut être impliqué dans différents modules de la maladie, ce qui signifie que différents modules de la maladie peuvent se chevaucher. Ces caractéristiques aident à l'identification du module de la maladie, une étape importante de la médecine des réseaux. » (Barabási *et al.,* 2011, p. 58–59)

(1) Tout d'abord, chaque maladie est identifiée par un unique module de la maladie. Cela signifie que chaque maladie-type peut être instanciée dans un module spécifique (par exemple, la tuberculose dans le module de la tuberculose). (2) Par ailleurs, si chaque module de la maladie est spécifique d'une maladie donnée, des modules de maladies différentes peuvent se chevaucher. En particulier, plus deux maladies sont proches phénotypiquement (en termes de symptômes, d'évolution, de localisation anatomoclinique, etc.) et plus les modules des deux maladies se chevaucheront : cela signifie qu'un même composant peut être impliqué dans plusieurs modules et c'est là l'explication de la comorbidité. (3) Enfin, pour parvenir à identifier un module de la maladie, il faut faire l'hypothèse initiale que module topologique, module fonctionnel et module de la maladie sont une même entité. C'est une exigence de la méthode expérimentale qui repose sur l'hypothèse locale établie

dans l'étude du diseasome (voir Figure 49 pour les méthodes d'identification d'un module fonctionnel). Mais en pratique, le module de la maladie finalement validé est rarement strictement superposable au module topologique et au module fonctionnel identifiés initialement.

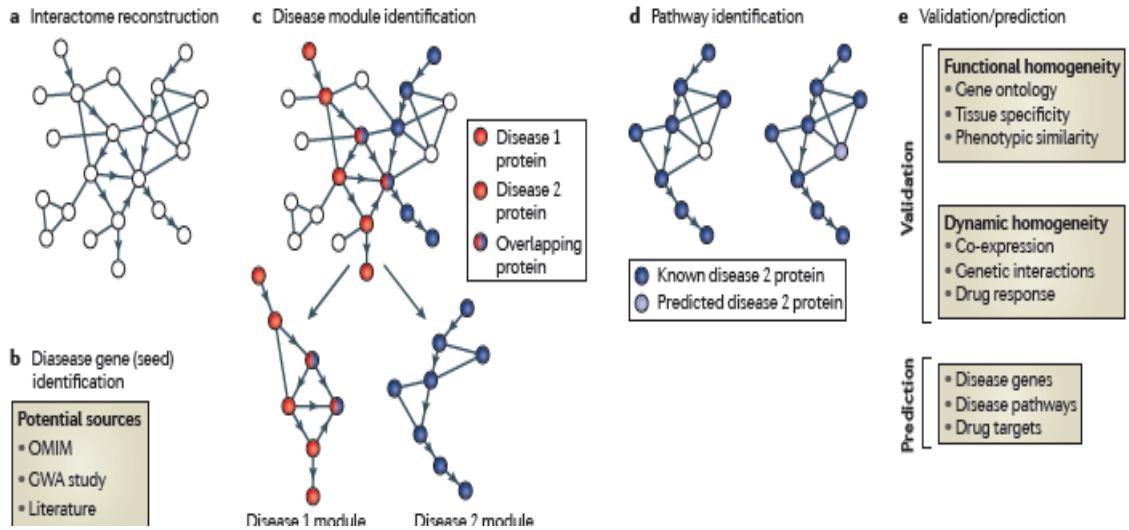


Figure 49 : Les étapes de l'identification et de la validation d'un module de la maladie (Barabási et al., 2011, p. 62). Pour une maladie donnée, l'identification et la validation d'un module de la maladie est un processus en plusieurs étapes : (a) Reconstruire l'interactome. (b) Identifier un ensemble de gènes-candidats potentiels. (c) Une fois que les gènes-candidats potentiels sont placés sur l'interactome, il s'agit d'identifier des sous-réseaux qui contiennent le maximum de composants associés à la maladie. Ce sous-réseau est alors considéré comme un module de la maladie potentiel. Plus deux maladies sont similaires dans leurs manifestations phénotypiques, plus on s'attend à ce que les modules associés aux deux maladies se chevauchent. (d) Lorsqu'un module potentiel contient un grand nombre d'éléments, on cherche à identifier des chemins métaboliques particuliers (*disease pathway identification*) dont la rupture pourrait être spécifiquement responsable du phénotype pathologique. On utilise alors le principe de parcimonie des réseaux pour sélectionner les chemins métaboliques les plus probables, en assumant que le chemin causal est le chemin le plus court pour connecter les composants de la maladie. (e) La validation des modules implique de tester leur homogénéité fonctionnelle et dynamique. La méthode de validation dépend des outils et des données disponibles : données sur l'expression des gènes, associations pangénomiques, etc. Finalement, on teste le module de la maladie en question sur des modèles animaux.

6.3.1.2. Mais chaque individu a un module de la maladie spécifique

La médecine des réseaux identifie donc chaque maladie-type à son module, mais comment passer du module de la maladie-type au module de la maladie instance ? En effet, du fait de la redondance fonctionnelle et de la robustesse des réseaux cellulaires, toute perturbation n'entraîne pas nécessairement l'apparition d'une maladie. La redondance fonctionnelle est une propriété bien établie des

systèmes biologiques, qui signifie, dans le langage de la biologie des systèmes, qu'une même fonction est généralement assurée par plusieurs chemins métaboliques distincts. Sans cette redondance fonctionnelle, les organismes vivants et les humains en particulier seraient peu robustes et perpétuellement malades, puisque la moindre perturbation d'un seul module fonctionnel serait suffisante pour entrainer l'apparition d'une maladie. Redondance fonctionnelle et robustesse sont donc intrinsèquement liées. Annick Lesne distingue différents types de perturbation d'un module de la maladie (voir Figure 50), soulignant qu'une perturbation du module n'entraîne pas systématiquement son inactivation : l'efficacité du module peut être simplement diminuée, ou le module peut acquérir une fonction différente, en fonction du type de perturbation observée (Debret *et al.,* 2011).

Cette distinction entre différents types de perturbations correspond bien aux différents types d'événements pouvant entraîner des mutations, aux différents types de mutations qu'un généticien identifie pour un allèle donné et aux différentes conséquences que ces mutations peuvent avoir sur le phénotype correspondant. Ainsi quand on qualifie une mutation, celle-ci peut avoir différentes origines (spontanée, induite, héritée), toucher différentes localisations (spontanée, germinale, séquence régulatrices, introns, exons, séquences 5' et 3' UTR, etc.), être de différents types (substitution ; délétion ; addition ; remaniement chromosomique ; insertion) et avoir différents effets sur le produit allélique : mutation silencieuse, perte de fonction totale (allèle amorphe), perte de fonction partielle (allèle hypomorphe), dominant négatif (le produit de l'allèle mutant inhibe le produit de l'allèle sauvage), gain de fonction (allèle hypermorphe avec augmentation du niveau d'expression du gène ou de l'activité du produit), nouvelle fonction (allèle néomorphe), mutation conditionnelle (par exemple, thermosensible), mutation à effet pléiotrope (responsable de plusieurs phénotypes).

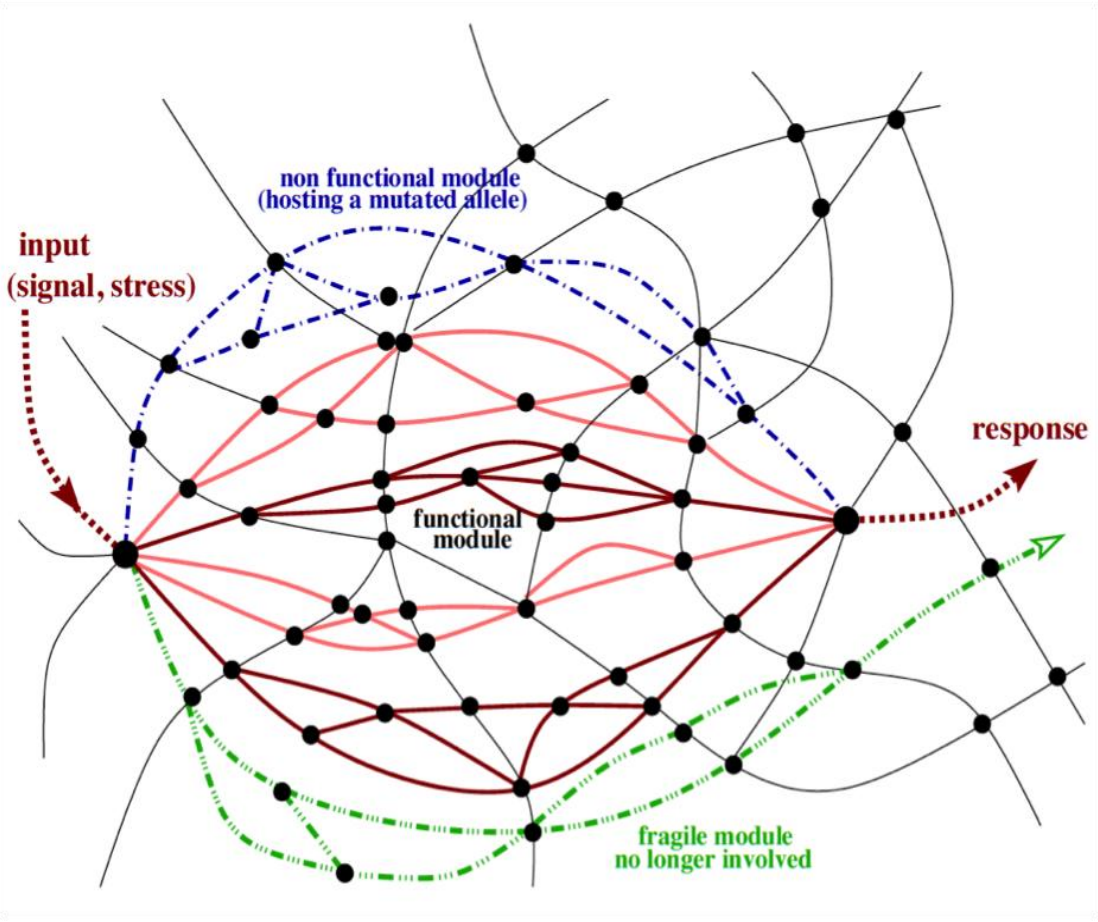


Figure 50 : Variété des perturbations possibles d'un module fonctionnel de la maladie *(Debret et al.,* 2011). Dans cette représentation, chaque arête soulignée en gras représente un module, chaque module est indépendant des autres et capable de produire une activité fonctionnelle ou une réponse propre. Parmi ces modules, certains (en vert) sont intrinsèquement non fonctionnels (bleus), parce qu'ils contiennent un allèle muté entraînant une inactivation de la fonction (c'est-à-dire une mutation allélique non sens ou faux sens). D'autres ne sont plus impliqués dans la réalisation de la fonction (en vert). Certains de leurs éléments sont parfois utilisés pour une autre fonction. Seuls les modules représentés par des lignes continues en gras sont des modules qui participent effectivement à la réalisation de la fonction. L'ensemble de ces modules fonctionnels définit le phénotype moléculaire d'un individu donné par rapport à la fonction considérée.

Pour comprendre le passage du module de la maladie-type au module de la maladie-instance et prendre en compte la diversité des perturbations d'un module de la maladie, Annick Lesne introduit donc le concept d'un « phénotype fonctionnel moléculaire idiosyncrasique et spécifique de l'âge ». Chaque individu a un ensemble de modules fonctionnels qui dépendent de son patrimoine génétique et de son histoire personne, c'est-à-dire d'événements contingents comprenant des mutations génétiques acquises au cours de la vie et des perturbations de l'environnement

intérieur (événements épigénétiques, événements stochastiques) et de l'environnement extérieur (polluants, nourriture, agents cancérigènes, etc.) :

> « Nous proposons la notion de « phénotype fonctionnel moléculaire » pour décrire le niveau intermédiaire entre le génotype et le phénotype observé, en fonction duquel un individu implémente une fonction et sa régulation dans un environnement donné. Il est spécifique de la fonction considérée. Il décrit comment un ensemble de modules est scindé pour un individu donné à un âge donné entre (i) des modules constitutivement déficients, typiquement parce qu'ils impliquent un allèle porteur d'une mutation délétère (ii) des modules qui ne sont plus utilisés pour la fonction considérée, typiquement parce que l'un de leurs éléments a été redirigé sur une autre fonction, et (iii) des modules qui sont véritablement recrutés pour la réalisation de sa fonction et de sa régulation. Ce phénotype moléculaire peut ainsi être disséqué et présenté au niveau des allèles impliqués (à travers le produit de leur expression ou à travers leurs séquences régulatrices) en des modules fonctionnels correspondants. Selon ce point de vue, les allèles sont soit (i) dans une forme non fonctionnelle (ii) soit dans une forme fonctionnelle mais non impliqués dans la fonction (iii) fonctionnels et véritablement impliqués dans la fonction. » (Debret *et al., 2011, p. 508)*

Pour passer de l'explication de la maladie-type à la maladie instance, il faut donc identifier précisément quelles sont les perturbations du module de la maladie-type présentes chez un individu donné, en prenant en compte l'ensemble de son histoire personnelle.

Ainsi, en s'appuyant sur une définition de la maladie en termes de perturbation d'un module fonctionnel, définition qui s'applique potentiellement à toutes les maladies, la médecine des réseaux est bien une théorie générale de la maladie, proche dans ses ambitions de la théorie hippocratico-galénique qui définissait la maladie en général en termes de déséquilibre entre quatre humeurs. Cette définition signifie que toute maladie-type peut être définie par un module fonctionnel de la maladie unique. Mais une maladie instance (un cas particulier de tuberculose par exemple) peut aussi être expliquée par l'identification spécifique du ou des composants du module dont l'efficacité fonctionnelle est diminuée, inactivée ou modifiée et qui dépend de

l'histoire personnelle d'un individu donné (âge, environnement, patrimoine génétique, etc.).

6.3.1.3. La médecine des réseaux, un cadre interactionniste co-constructionniste

Nous avons établi que la médecine des réseaux était une théorie générale de la maladie, mais cette théorie générale de la maladie est aussi ancrée dans un interactionnisme co-constructionniste. Considérons la composition d'un module de la maladie. Un module de la maladie n'a pas pour unique composant des gènes, il comprend aussi des facteurs de transcription, des protéines, des cascades de réactions biochimiques, etc. C'est d'ailleurs pourquoi la médecine des réseaux ne s'intéresse pas seulement aux réseaux gènes-maladies, mais aux réseaux gènes-facteurs de transcription, aux réseaux gènes-gènes, aux réseaux gènes-protéines, aux réseaux protéines-protéines, etc. La notion de module de la maladie n'est d'ailleurs pas qu'un concept « interne » à l'organisme : il prend en compte les déterminants de l'environnement.

Ainsi, dans un article de 2007, publié par les mêmes auteurs que l'article sur le diseasome, le module de la malade est décomposé entre quatre réseaux modulaires différents qui interagissent pour produire le phénotype final : (a) le génome primaire de la maladie G est l'ensemble des anomalies moléculaires, qu'elles soient héréditaires ou acquises, qui ont été associées au phénotype général, (b) le génome secondaire de la maladie D comprend l'ensemble des gènes modificateurs de la maladie et leurs polymorphismes ou haplotypes, (c) le phénotype intermédiaire I comprend l'ensemble des polymorphismes qui influencent chacune des réponses génériques de l'organisme au stress, (d) le module E comprend les déterminants environnementaux qui *« interagissent avec les différents sous-génomes pour modifier la transcription des gènes qui les composent et pour moduler la traduction de leurs produits protéiques et leurs modifications post-traductionnelles, entrainant ainsi des changement dans le métabolisme des protéines et dans leurs fonctions cellulaires et définissant un phénotype intermédiaire »* (Loscalzo *et al., 2007, p. 3)*. Les interactions entre ces quatre types de réseaux permettent de définir les états physiopathologiques (PS) qui, à leur tour, déterminent l'ensemble des phénotypes pathologiques P, c'est-à-dire

l'ensemble des phénotypes qu'une maladie-type peut prendre, c'est-à-dire l'ensemble des maladies-instances (voir Figure 51). Précisons que le terme d'état physiopathologique (PS) désigne l'ensemble des perturbations infra-cliniques qui sont responsables des symptômes variés des cas de maladie-instance. Ce sont donc bien des interactions bidirectionnelles (comme on le voit sur la Figure 51), qui sont envisagées entre ces différents modules de la maladie, dessinant une matrice causale complexe interactionniste et co-constructionniste de la maladie.

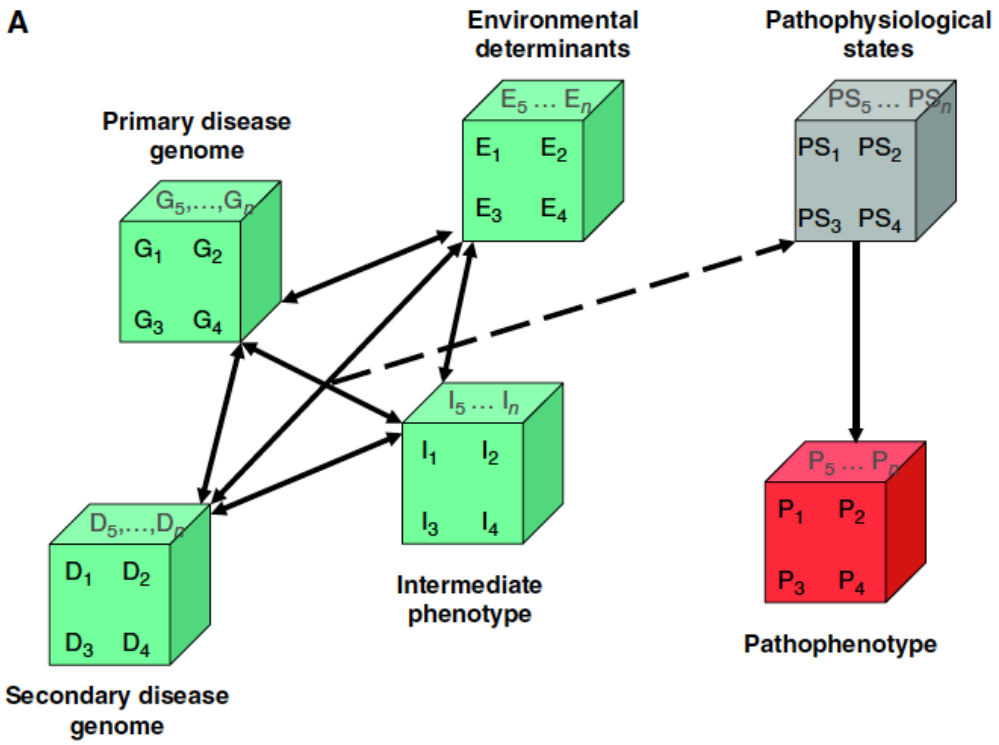


Figure 51 : Combinaison de quatre réseaux modulaires impliqués dans le développement d'une maladie (Loscalzo *et al.*, 2007, p. 4). Quatre réseaux modulaires (le génome primaire, le génome secondaire, le phénotype intermédiaire, les déterminants environnementaux) interagissent (de façon co-constructionniste, comme le suggèrent les arêtes bi-directionnelles) pour conduire à l'apparition d'états physiopathologiques (les symptômes). Plusieurs combinaisons de symptômes sont possibles, qui donnent lieu à plusieurs types de pathophénotypes, c'est-à-dire à une variabilité clinique interindividuelle dans la présentation de la maladie.

Prenons un exemple plus précis : la Figure 52 représente l'ensemble des réseaux modulaires impliqués dans la drépanocytose. Le gène HbS (rose) représente bien sûr le génome primaire de la maladie. Différents gènes (en gris) dont on a rappelé précédemment qu'ils étaient des gènes modificateurs, susceptibles de modifier

l'expression du gène Hbs (gènes associés à d'autres anomalies des globules rouges, etc.) représentent le génome secondaire de la maladie. En jaune sont représentés les phénotypes intermédiaires, c'est-à-dire un ensemble de phénomènes cellulaires qui sont mis en jeu lorsque l'organisme d'un patient atteint de la drépanocytose est soumis à des conditions environnementales particulières : réponse auto-immune, inflammation, etc. En vert sont représentés les déterminants environnementaux dont a déjà discuté l'importance dans l'apparition des symptômes d'une maladie donnée. Enfin, en bleu apparaissent les états physiopathologiques intermédiaires qui sont l'ensemble des symptômes possibles de la maladie.

Cette matrice causale témoigne d'un interactionnisme co-constructionniste et est donc un formidable outil de compréhension générale de la maladie-type et d'instanciation des cas particuliers. Du point de vue de la compréhension de la physiopathologie de la maladie-type, on remarquera en particulier la possibilité de comprendre les boucles de rétroaction : par exemple, les patients drépanocytaires qui ont une mutation homozygote dans le gène HbS ont une réponse auto-immune diminuée qui les rend plus susceptibles aux infections. Mais à cause de la fièvre, l'infection entraine une hypoxie et une déshydratation, qui sont elles-mêmes la cause de crises vaso-occlusives, etc. Du point de vue de la compréhension de la physiopathologie des maladies instances, une telle représentation permet aussi bien de comprendre le cas de l'hétérozygote asymptomatique dans des conditions environnementales normales, mais qui déclenche des crises aigues lorsqu'il est confronté à certaines conditions particulières (situation d'hypoxie en altitude par exemple), que l'anémie hémolytique chronique d'un patient homozygote.

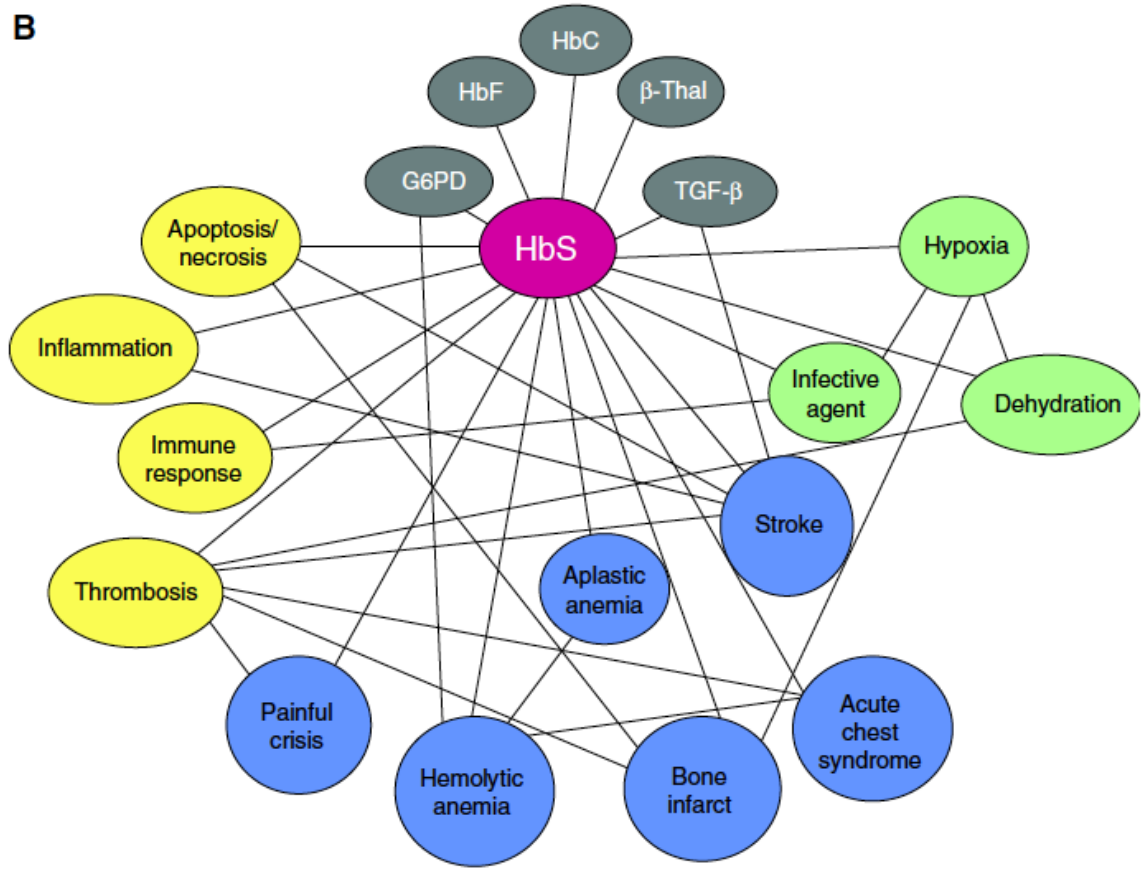


Figure 52 : Module fonctionnel de la drépanocytose (Loscalzo *et al.*, 2007, p. 4*)*. En rose : anomalie génétique principale ou génome primaire de la maladie G. En gris : gènes modificateurs de la maladie ou génome secondaire de la maladie D. En jaune : phénotypes intermédiaires I. En vert : déterminants environnementaux E. En bleu : états physiopathologiques ou symptômes de la maladie P.

6.3.1.4. La médecine des réseaux est-elle une théorie générale qui explique les quatre cibles de la maladie ?

Enfin, la médecine des réseaux est une théorie générale de la maladie car la définition modulaire de la maladie rend compte des quatre cibles de l'explication de la maladie.

(1) Les causes de la maladies : comme on le voit sur la Figure 51 et comme on l'a analysé pour l'exemple de la drépanocytose (Figure 52), la définition modulaire de la maladie permet de considérer à la fois l'ensemble des causes pouvant conduire à l'apparition d'une maladie-type et de distinguer les causes particulières d'une maladie-instance. Quatre grands types de causes possibles sont distingués : le génome primaire ou anomalie génétique principale G, le

génome secondaire ou gènes modificateurs de la maladie D, les déterminants environnementaux E et les états physiopathologiques intermédiaires. Un point intéressant à souligner est que cette conception permet aussi bien de rendre compte de l'action isolée d'une seule cause sur le phénotype pathologique final composé de l'ensemble des symptômes P, que de proposer une vision dynamique de l'enchaînement causal qui peut faire intervenir des causes de niveaux différents et qui peut comprendre des boucles de rétroaction (effet des symptômes sur l'expression du génome primaire, entrainant de nouveaux symptômes, etc.).

(2) Différents symptômes de la maladie peuvent recevoir des explications distinctes aussi bien du point de vue de la maladie-type que de la maladie-instance. Soulignons encore une fois que dans une représentation modulaire de la maladie, les symptômes agissent eux-mêmes comme une cause susceptible de modifier le module fonctionnel.

(3) L'évolution de la maladie : l'évolution de la maladie est expliquée par l'interaction dynamique des causes de différents niveaux. Notons également que la notion de phénotype fonctionnel moléculaire dépendant de l'âge permet également de prendre en compte à la fois l'évolution des modules fonctionnels d'une même maladie pour un individu au cours de sa vie et l'évolution conjointe de ces symptômes. Ce type de description paraît particulièrement adapté pour l'explication des maladies chroniques, comme la drépanocytose ou le diabète.

(4) Le traitement : c'est un point qui explique en grande partie l'attraction exercée par la médecine des réseaux (Antman *et al.,* 2012 *;* Jacunski et Tatonetti, 2013 *;* Nacher et Schwartz, 2012 ; Silverman et Loscalzo, 2013). En effet, l'innovation technologique en pharmacologie a considérablement diminué ces dernières années, la plupart des médicaments mis sur le marché sont ainsi des « me-too drugs », parfois traduits « médicaments d'imitation », qui reproduisent l'action d'un médicament déjà connu. Par ailleurs, on estime que les 1 200

médicaments approuvés par l'Agence américaine des produits alimentaires et médicamenteux (*Food and Drug Administration* ou FDA) ne ciblent qu'une faible partie de l'interactome (400 à 600 des 100 000 protéines estimées de l'interactome) et de surcroît qu'ils ciblent rarement des protéines directement impliquées dans la maladie, mais plus souvent des protéines voisines. En outre, le mythe de la « *magic bullet* », le médicament unique, spécifique d'une maladie donnée et qui suffirait à renverser complètement le cours d'une maladie est dépassé, tandis que la polypharmacologie qui combine plusieurs médicaments tend à se développer pour des maladies aussi différentes que le cancer, le diabète ou le VIH. Ce développement de la polypharmacologie ne va pas sans difficultés, étant donné l'importance considérable des effets secondaires qu'il a longtemps été difficile d'anticiper.

L'approche traditionnelle de la pharmacologie consiste à tester l'effet d'un médicament sur une maladie donnée, en surveillant les effets secondaires. La médecine des réseaux commence par caractériser d'une part le module fonctionnel de la maladie et d'autre part l'action pharmacologique du médicament. L'étape suivante consiste alors d'une part à identifier les propriétés émergentes et souvent non anticipées de la réponse du système au médicament et d'autre part à perturber systématiquement le réseau pour identifier les nœuds les plus importants pour l'action médicamenteuse. En renouvelant ce processus, il serait possible de développer une polypharmacie rationnelle, c'est-à-dire une combinaison de médicaments multiples qui maximise l'efficacité de la polypharmacie et qui minimise les effets secondaires, que ceux-ci soient dus aux interactions entre médicaments ou aux interactions du médicament avec d'autres cibles que les cibles initiales du médicament (Voir Figure 53).

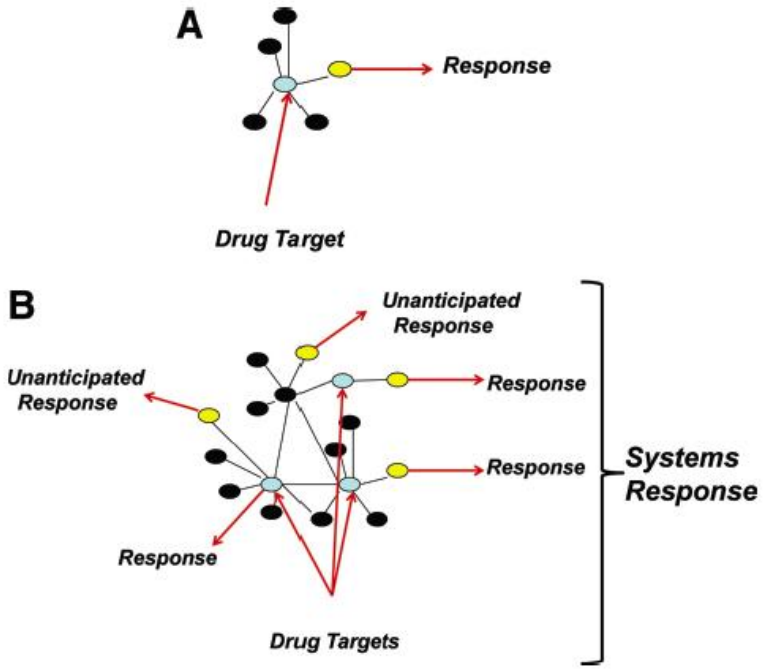


Figure 53 : Comparaison de l'approche pharmacologique conventionnelle et de l'approche pharmacologique fondée sur la médecine des réseaux (Loscalzo, 2012). A : dans une vision conventionnelle de la pharmacologie, on analyse une réponse unique de la cible du médicament (en bleu). B : dans une vision fondée sur les réseaux de la pharmacologie, la réponse monitorée est la réponse globale du système qui combine les réponses de plusieurs cibles d'un médicament.

6.3.2. La théorie génétique de la maladie, une théorie partielle au sein de la médecine des réseaux

Au sein de cette théorie générale, il nous semble qu'à partir de l'analyse du diseasome, on peut isoler une théorie partielle centrée sur le rôle commun des gènes dans toutes les maladies, autrement dit une théorie génétique de la maladie qui explique l'origine génétique commune des maladies et la distinction entre maladies monogéniques et polygéniques et qui permet de prédire de nouveaux gènes de la maladie.

6.3.2.1. Expliquer l'origine génétique commune des maladies

En effet, l'analyse globale du diseasome a montré que le réseau des maladies humaines était un réseau invariant d'échelle : la majorité des maladies sont connectées entre elles en un seul immense réseau parce qu'elles partagent des gènes communs, quelques maladies ont un degré très élevé et forment des *hubs* dans le réseau, la majorité des maladies sont peu connectées et certains groupes de maladies

très connectées entre elles correspondent à des classes de maladie. Cette propriété topologique du réseau suggère une origine génétique commune de nombreuses maladies, c'est-à-dire la façon dont les maladies sont des entités interdépendantes d'un point de vue génétique. Cette origine génétique commune permet de rendre compte de trois classes de faits particuliers : l'explication de la comorbidité, l'explication des familles de syndromes et l'explication des classes de maladie.

La comorbidité désigne le fait de présenter plusieurs maladies, en dehors de la maladie primaire, de façon concomitante. Un des cas les mieux explorés de comorbidité dans les maladies somatiques est celui de l'obésité et du diabète. L'hypothèse de l'origine génétique commune des maladies permet de donner une explication de ce phénomène : c'est parce que deux maladies A et B partagent un ou plusieurs gènes qu'avoir la maladie A augmente la probabilité de développer la maladie B : une mutation d'un gène du module fonctionnel de la maladie A a de grandes chances d'avoir des conséquences fonctionnelles sur le module de la maladie B, si les modules de la maladies A et B ont ce gène en commun. Cette hypothèse a été testée : une étude a ainsi estimé qu'en moyenne, un patient avait deux fois plus de chances de développer une maladie particulière si cette maladie partage un gène avec la maladie primaire développée par le patient. Cependant, quand on analyse individuellement chaque paire de maladies connectées, certaines paires qui partagent des gènes ne montrent pas de risque plus important de comorbidité. Cette observation ne remet pas en cause l'explication de la comorbidité, elle doit être comprise dans le cadre interactionniste d'une explication de la maladie qui analyse le rôle causal d'un gène dans son contexte (i.e. en fonction du rôle qu'il joue dans un module fonctionnel donné) : étant donné le phénomène de pléiotropie selon lequel différentes mutations dans un même gène peuvent avoir différentes conséquences sur le produit du gène, certaines mutations d'un même gène peuvent être spécifiques d'une maladie particulière et ne pas avoir de conséquences sur le développement de maladies qui sont causées par le même gène mais par des mutations différentes. Il a d'ailleurs été démontré que deux maladies sont causées par des mutations d'un même gène qui affectent le même domaine fonctionnel d'une protéine (et a donc des conséquences fonctionnelles similaires) ont un risque de comorbidité plus élevé que

des maladies qui sont causées par des mutations d'un même gène mais qui affectent des domaines fonctionnels distincts de la même protéine (Park *et al., 2009)*.

Un syndrome désigne généralement une maladie qui n'a pas de cause identifiée, présentant un ensemble de symptômes variés sans lien apparent les uns avec les autres. Les familles de syndromes sont des groupes de maladies qui semblent Avoir des symptômes en commun mais dont la cause principale n'est pas comprise. C'est à nouveau l'origine génétique commune de ces syndromes qui permet de comprendre l'existence de familles de syndromes : ces syndromes distincts partagent des caractéristiques phénotypiques ou symptômes similaires, parce qu'ils partagent des gènes fonctionnellement proches, c'est-à-dire qui codent des produits protéiques impliqués dans la réalisation de mêmes fonctions ou de fonctions proches. Ainsi, dans la Figure 54, trois syndromes (le syndrome Stickler, le syndrome Marshall et le syndrome OSMED) ont des symptômes communs. Ces syndromes sont causés par des mutations dans des gènes fonctionnellement reliés : COL2A1, COL11A1 et COL11A2. Le syndrome Pallister-Hall est distinct, il est causé par des mutations dans le gène GLI3 qui est fonctionnellement non relié. Plusieurs gènes peuvent entrainer un seul phénotype, comme c'est le cas dans le syndrome Stickler qui peut être causé par des mutations dans chacun des trois gènes. À l'inverse, un gène peut causer différents syndromes, comme c'est le cas de COL11A1 (responsable des syndromes Stickler et OSMED) et de COL11A2 (responsable des syndromes Stickler et OSMED). L'épaisseur des traits noirs qui relient les gènes entre eux représente leur degré de parenté fonctionnelle hypothétique.

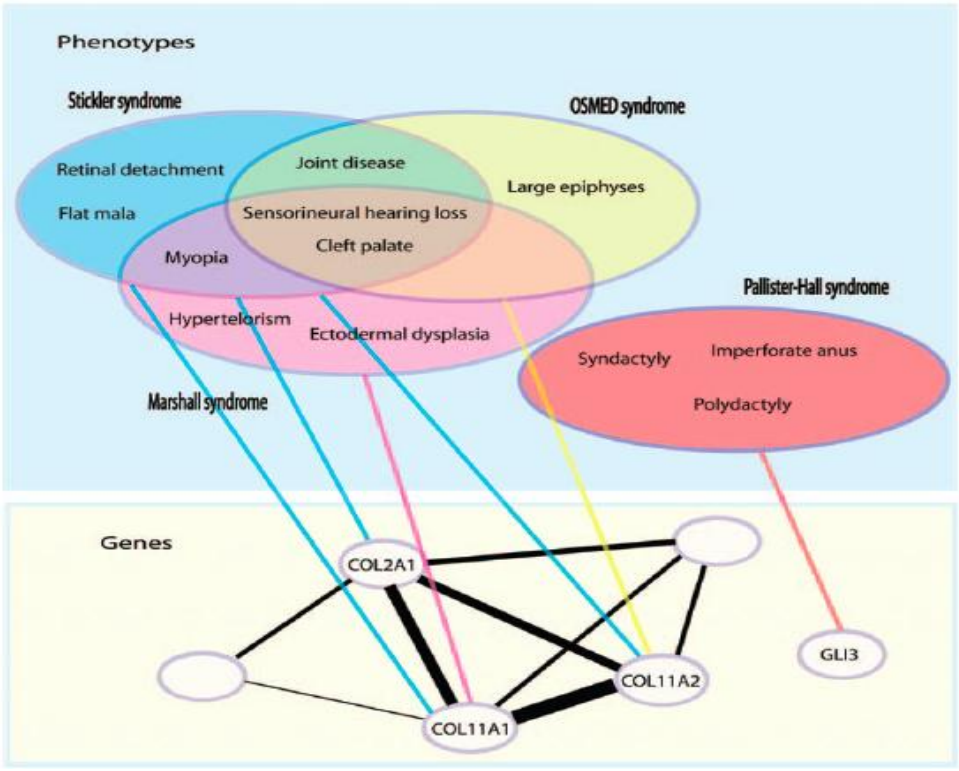


Figure 54 : Relations possibles entre des gènes et des phénotypes dans une famille de syndromes (Oti et Brunner, 2006, p. 2). Les symptômes communs du syndrome Stickler, du syndrome Marshall et du syndrome OSMED sont expliqués par une origine génétique commune. Le syndrome Pallister-Hall, qui en revanche ne partage aucun gène commun avec ces trois syndromes, entraine des symptômes distincts.

Enfin, certaines classes de maladie (celles qui forment des groupes de maladies très connectées dans le diseasome) peuvent aussi être comprises comme ayant une origine fonctionnelle commune. Ainsi, le fait que les cancers forment un grand groupe hyperconnecté au sein du diseasome est parfaitement cohérent avec les observations selon lesquelles l'apparition du cancer met en jeu des mutations dans un petit nombre de gènes, les oncogènes et les anti-oncogènes qui contrôlent de grandes fonctions cellulaires (comme l'apoptose, la réparation de l'ADN, etc.). Pour d'autres classes de maladie, comme les classes fondées sur une localisation anatomoclinique, on observe ce même phénomène, qui est en cohérence avec l'hypothèse locale que nous avons analysée dans le 6.2.3 et selon laquelle les gènes qui sont impliqués dans une même maladie codent des protéines qui sont coexprimées de façon synchronisée et qui interagissent dans les mêmes tissus. Cependant, pour les classes de maladie qui sont plus difficiles à délimiter sur le plan anatomoclinique (les maladies neurologiques ou

les maladies métaboliques par exemple), les maladies sont très dispersées dans le réseau des maladies humaines. Par ailleurs, certaines associations sont des plus surprenantes, comme le fait que la maladie d'Alzheimer et les maladies cardiovasculaires sont fortement connectées.

On entrevoit ici une possible conséquence de la médecine des réseaux : dans la mesure où certaines classes de maladie ne semblent pas avoir d'origine génétique commune, on pourrait envisager de modifier au moins partiellement nos classifications de la maladie et/ou la façon dont les maladies sont délimitées entre elles.

6.3.2.2. Redéfinir la distinction entre maladies monogéniques et maladies polygéniques

Notons tout d'abord qu'une première façon de redéfinir la distinction entre maladies monogéniques et maladies polygéniques a été proposée dès les premières analyses du diseasome (voir Figure 55), en montrant comment différentes combinaisons entre les quatre grandes causes de la maladie (génome primaire, génome secondaire, phénotypes physiopathologiques intermédiaires, déterminants environnementaux) correspondent à différents types de maladies.

Cette représentation a des avantages indéniables puisqu'elle permet d'intégrer les phénomènes de pléiotropie, d'hétérogénéité génétique, d'hétérogénéité allélique et de gènes modificateurs de la maladie, qui déstabilisent le concept de maladie monogénique et qui ont contribué à effacer la distinction traditionnelle entre maladie monogénique et maladie polygénique. Notons d'ailleurs que les auteurs ne donnent pas d'exemple de maladie monogénique au sens le plus strict du terme et que leur exemple de maladie environnementale est discutable : comme on l'a vu dans le chapitre 5, on pourrait imaginer des gènes de prédisposition ou de résistance à l'endocardite infectieuse aiguë, qui est une infection des valves du cœur.

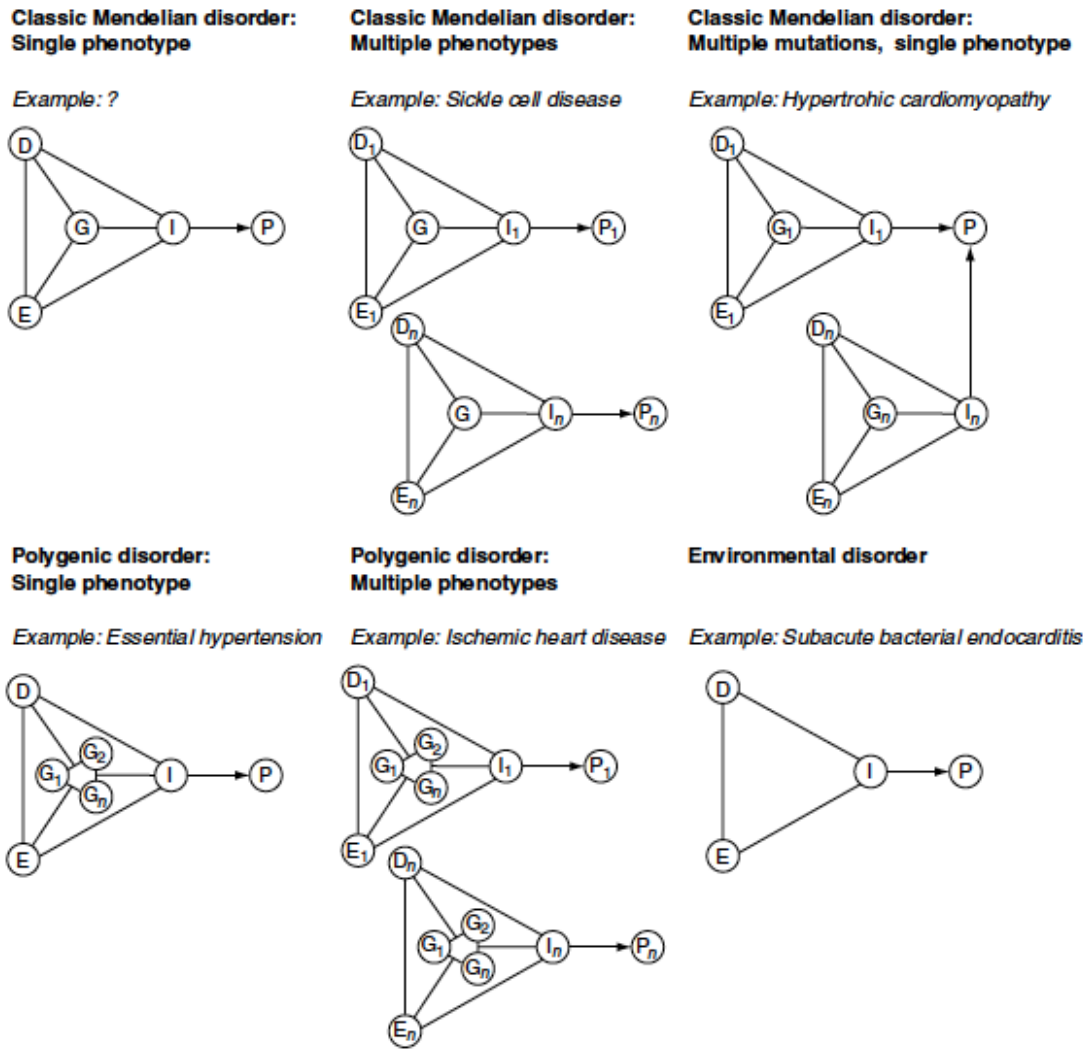


Figure 55 : Représenter la variété des maladies monogéniques et polygéniques à partir d'une définition modulaire de la maladie (Loscalzo *et al.*, 2007). En utilisant les quatre réseaux modulaires de la maladie (génome primaire, génome secondaire, phénotype intermédiaire et déterminants environnementaux), il est possible de représenter le spectre des relations complexes entre génotype et phénotype dans l'explication des maladies, des maladies « les plus monogéniques » aux maladies « les plus environnementales ». Ce type de représentation permet ainsi de rendre compte des phénomènes d'hétérogénéité génétique et de pléiotropie (2e figure en haut), d'hétérogénéité allélique (3e figure en haut) et des différentes formes d'interactions gène x gène et gène x environnement. On remarquera que les auteurs ne donnent pas d'exemple de maladie mendélienne classique, ce qui confirme bien le bouleversement du concept de maladie génétique.

À l'instar d'Annick Lesne, on peut cependant faire un pas supplémentaire, et redéfinir cette distinction en s'appuyant à la fois sur le concept de phénotype moléculaire idiosyncrasique et sur les propriétés topologiques du diseasome. En effet, le fait que le diseasome soit un réseau invariant d'échelle a une deuxième conséquence : la spécificité de sa robustesse. Comme on l'a vu dans la section 6.1.3 de ce chapitre, les réseaux invariants d'échelle sont associés à une certaine forme de robustesse ; ils sont robustes aux attaques aléatoires et sensibles aux attaques

spécifiques sur les gènes qui forment des *hubs* au sein du réseau. Par ailleurs, robustesse et redondance fonctionnelle sont liées. Plus un module fonctionnel de la maladie admet des sous-modules indépendants susceptibles de prendre le relais de la fonction biologique considérée en cas de défaillance des sous-modules impliqués, plus le système est robuste.

À partir de ces deux propriétés, on peut redéfinir les maladies monogéniques comme des maladies dont la redondance fonctionnelle est faible et dont par conséquent la robustesse est faible ; tandis que les maladies polygéniques sont comprises comme des maladies à redondance fonctionnelle élevée et dont par conséquent la robustesse est élevée. Pour le dire différemment, si le module fonctionnel d'une maladie donnée ne présente qu'un seul chemin métabolique pour réaliser la fonction correspondante, la moindre perturbation (par exemple une seule mutation génétique) suffit à désorganiser suffisamment le module pour atteindre le seuil où apparaît la maladie. Si au contraire le module fonctionnel d'une maladie est composé de plusieurs sous-modules redondants, cela signifie qu'il est nécessaire que plusieurs événements concourent à l'inactivation de plusieurs sous-modules pour atteindre le seuil d'apparition de la maladie (voir Tableau 9).

Cette redéfinition de la distinction entre maladie monogénique et maladie polygénique n'aboutit pas à constituer deux nouvelles catégories de maladies (maladies hautement redondantes et robustes d'une part et maladies faiblement redondantes et robustes d'autre part), mais elle a l'avantage de s'éloigner définitivement de tout génocentrisme, tout en permettant de rendre compte de la diversité des modules fonctionnels des maladies. Enfin, cette redéfinition est tout à fait cohérente avec l'observation générale, selon laquelle les maladies considérées plutôt comme monogéniques ont tendance à se déclarer plus tôt dans la vie d'un individu que les maladies dites polygéniques : c'est qu'une seule perturbation est susceptible d'aboutir à l'inactivation du module fonctionnel de la maladie dans le premier cas, tandis qu'une série d'événements est nécessaire dans le second.

	Génétique classique	Théorie génétique dans la médecine des réseaux
Maladies monogéniques	Une mutation mendélienne héritée d'un gène > une protéine non fonctionnelle > une maladie	Redondance fonctionnelle faible, Robustesse faible > une mutation peut suffire à déclencher la maladie
Maladies polygéniques	De multiples mutations dans de multiples gènes + plusieurs facteurs environnementaux > maladie	Redondance fonctionnelle élevée Robustesse élevée > plusieurs événements sont nécessaires pour déclencher la maladie
Distinction maladies monogéniques/ maladies polygéniques	Continuum génétique Les maladies monogéniques sont « plus génétiques » que d'autres	Continuum en termes de redondance et de robustesse

Tableau 9 - La distinction entre maladies monogéniques et maladies polygéniques dans la génétique classique et dans la théorie génétique de la médecine des réseaux

6.3.2.3. Prédire de nouveaux gènes de la maladie

Il nous faut enfin souligner le rôle hautement heuristique de la médecine des réseaux : en utilisant les propriétés topologiques du diseasome, la médecine des réseaux permettrait de prédire de nouveaux gènes de la maladie ou plus exactement d'identifier des gènes candidats pour une maladie donnée. Il existe trois techniques de référence (voir Figure 56). Dans la méthode de liaison, les gènes qui interagissent avec des candidats connus (interactions directes ou voisinage immédiat) sont considérés comme des candidats possibles. Dans la méthode modulaire, on identifie d'abord le module de la maladie et on considère que les gènes dont les produits interagissent avec ce module sont des candidats possibles. Dans la méthode de diffusion enfin, on explore par algorithme le voisinage d'un module de la maladie. Cette méthode est fondée sur l'idée qu'il existe dans un réseau des nœuds « goulots d'étranglement » qui font le lien entre deux modules A et B et qui potentiellement permettent la diffusion d'une perturbation d'un module A à un module B.

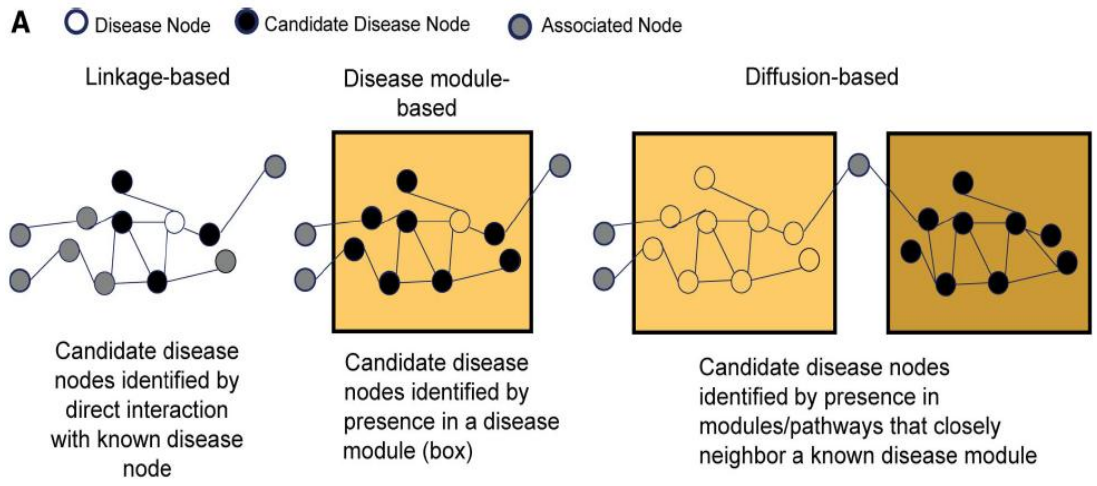

A
Disease Node
Candidate Disease Node
Associated Node
Linkage-based
Disease module-based
Diffusion-based
Candidate disease nodes identified by direct interaction with known disease node
Candidate disease nodes identified by presence in a disease module (box)
Candidate disease nodes identified by presence in modules/pathways that closely neighbor a known disease module

Figure 56 : Trois méthodes pour prédire les gènes candidats d'une maladie : méthode de liaison, méthode fondée sur le module de la maladie et méthode de diffusion (Chan et Loscalzo, 2012). Méthode de liaison (à gauche) : les gènes situés dans l'intervalle de liaison d'une maladie dont les protéines interagissent avec une protéine connue associée à la maladie, sont considérés comme des gènes candidats potentiels de la maladie. Méthode fondée sur le module de la maladie : des algorithmes sont utilisés pour regrouper des gènes fortement interconnectés, dans l'espoir de mettre à jour des modules fonctionnels potentiels de la maladie dans l'interactome. Les gènes qui font partie de ces regroupements sont considérés comme des gènes candidats potentiels. Méthode de diffusion : on utilise des algorithmes ou des informations fonctionnelles pour identifier des gènes appartenant à un module voisin du module d'intérêt. Par exemple, si deux modules sont impliqués dans la même réaction métabolique par une protéine en commun, on considère que les gènes du module voisin sont des gènes-candidats potentiels.

En conclusion, au sein de la médecine des réseaux, comprise comme une théorie générale interactionniste co-constructionniste de la maladie, qui repose sur une définition modulaire de la maladie, et qui explique les causes, les symptômes, l'évolution et le traitement des maladies, nous avons identifié une théorie génétique de la maladie. Cette théorie génétique explique l'origine génétique commune des maladies, renouvelle la distinction traditionnelle entre maladies monogéniques et maladies polygéniques et a un fort pouvoir heuristique de prédiction de nouveaux gènes candidats de la maladie.

6.3.3. Quel type d'explication mobilise la médecine des réseaux ?

Il s'agit à présent de déterminer quel type d'explication la médecine des réseaux et sa théorie génétique de la maladie mobilisent. Dans cette section, nous caractériserons d'abord la médecine des réseaux comme une explication topologique où l'explication d'un système dérive des propriétés topologiques de ces parties. La

médecine des réseaux ne saurait cependant se résumer à une explication topologique, car d'une part, elle fait constamment appel à des informations mécanistes et d'autre part, elle incite à trouver des explications mécanistes qui permettent de comprendre l'origine des propriétés topologiques qu'elle manifeste.

6.3.3.1. Explication topologique

Nous avons insisté sur l'existence de trois formes principales (mais non exhaustives) d'explication en médecine que sont l'explication physiopathologique ou explication mécaniste, l'explication évolutionnaire et l'explication épidémiologique. Nous avons montré cependant que les trois classes de faits spécifiques d'une théorie génétique de la médecine des réseaux s'appuient sur les propriétés topologiques du diseasome. Il s'agit donc ici d'une autre forme d'explication que Philippe Huneman appelle « explication topologique » et que l'on peut également retrouver en écologie (Montoya *et al., 2006)* ou dans les neurosciences (Sporns, 2012).

Philippe Huneman définit l'explication topologique comme *« une forme d'explication qui fait abstraction des relations causales et des interactions dans un système, dans le but de détecter un certain type de propriétés « topologiques » d'un système et de tirer de ces propriétés des conséquences mathématiques qui expliquent les caractéristiques du système visé. »* (Huneman, 2010, p. 214). Outre cette définition, Philippe Huneman propose une caractérisation formelle subtile de l'explication topologique et des propriétés topologiques, dont nous nous dispenserons cependant, car elle n'est pas centrale pour notre propos. Il est en revanche assez simple de comprendre la nature de l'explication topologique, en la comparant à l'explication mécaniste. Au contraire des explications mécanistes, les explications topologiques ne sont pas fondées sur les propriétés physiques ou matérielles des parties d'un système, mais sur leurs propriétés topologiques représentées dans un espace abstrait :

> « Quand nous expliquons un phénomène, nous considérons généralement comment le système, dont ce phénomène est une propriété, un résultat, une caractéristique, ou une conséquence, se comporte. Nous pouvons essayer de

> déterminer des relations causales, ou de subsumer différentes caractéristiques de ce système et de son fonctionnement sous certaines lois de la nature – ces deux ambitions n'étant pas mutuellement exclusives. Le système considéré peut avoir de nombreuses propriétés. Certaines d'entre elles concernent la façon dont, pour le dire grossièrement, le système remplit un espace, la façon dont les parties du système sont situées dans l'espace l'une par rapport à l'autre et si ces relations demeurent constantes sous certaines déformations continues du système (et si oui, lesquelles). C'est cela que j'appelle « les propriétés topologiques » du système. » (Huneman, 2010, p. 214)

Ainsi, pour obtenir une explication topologique, il faut d'abord définir un système (par exemple un écosystème d'espèces), prendre les parties de ce système (en l'occurrence, les espèces) et les représenter en faisant abstraction de la nature de leurs propriétés physiques (par exemple, par des points dans un espace). Dans un second temps, il s'agit de représenter une propriété relationnelle (par exemple la relation proie-prédateur) de ce système : on relie alors par des arêtes fléchées les espèces qui sont des prédatrices à leurs proies, à partir de données expérimentales précédemment recueillies. Une fois que ces informations sont représentées dans un espace qui fait abstraction de toutes les autres propriétés du système (par exemple les comportements de coopération entre certaines espèces), il s'agit d'utiliser la façon dont les parties du système (en l'occurrence les espèces) sont représentées dans l'espace pour expliquer le comportement de tout le système (l'ensemble des nœuds) ou d'un segment du système (un groupe de nœuds en particulier). Cet exemple choisi par Philippe Huneman aboutit à la représentation d'une communauté écologique dans un parc où les nœuds sont les espèces et où les arêtes fléchées représentent les relations proie-prédateur (voir Figure 57).

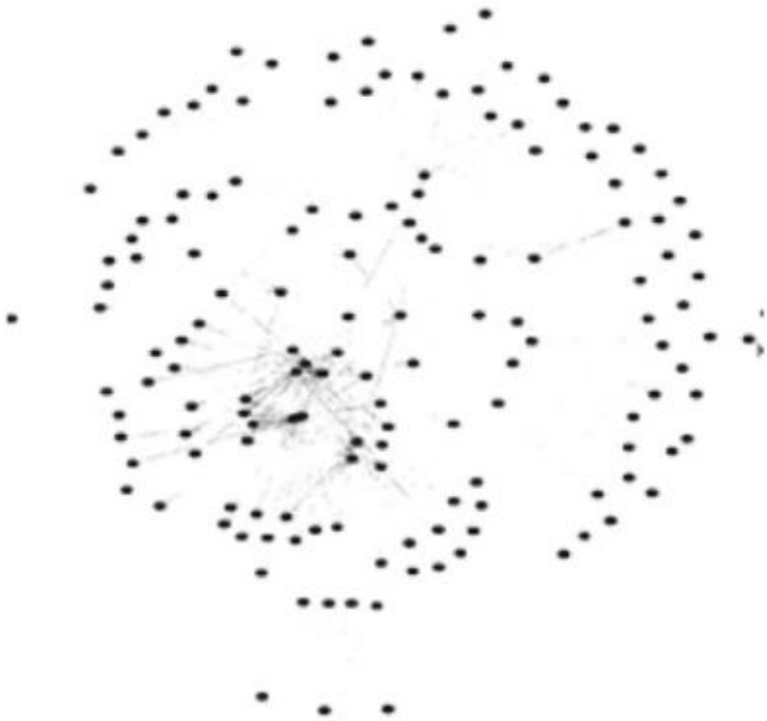


Figure 57 : Un réseau écologique dans une communauté (Huneman, 2010)

C'est ce même type d'explication topologique que l'on retrouve avec la médecine des réseaux et avec l'analyse du diseasome, mais est-ce le seul type d'explication ?

6.3.3.2. Explication mécaniste

Nous avons défini dans un premier temps l'explication topologique en la distinguant des explications mécanistes. Mais cette distinction n'est pas absolue, en fonction de la façon dont on définit les explications mécanistes : c'est ainsi que Woodward qualifie les explications topologiques d'explications « relativement » non-mécanistes. Pour Woodward, ces deux formes d'explication ne sont d'ailleurs pas en compétition mais se complètent mutuellement, car les explications topologiques utilisent souvent des informations mécanistes.

> « […] j'ai tendance à penser que l'image selon laquelle ces deux variétés d'explication peuvent se renforcer ou être complémentaires l'une de l'autre, est souvent correcte. Par exemple, les explications (relativement) non

> mécanistes qui font appel à la structure des réseaux […] peuvent néanmoins être construites à partir d'information mécaniste, au sens où le réseau lui-même est une représentation abstraite de faits à propos de relations mécanistes, même si ce sont les caractéristiques générales de la topologie des réseaux et non le détail de ces relations mécanistes qui, si l'on peut dire, dirigent les explications fournies. » (Woodward, 2013, p. 41)

En effet, la médecine des réseaux utilise sans cesse ce type d'informations mécanistes ou tout du moins, elle présuppose que les relations qu'elles représentent sont sous-tendues par des mécanismes. Par exemple, la construction du diseasome est fondée sur la base de données OMIM qui répertorie les associations entre un variant allélique et une maladie donnée. Certaines de ces données sont le résultat d'associations pangénomiques, c'est-à-dire sont des associations statistiques, mais si ces associations sont considérées avec confiance c'est bien parce que l'on présuppose qu'il existe un mécanisme sous-jacent qui relie le variant allélique à la maladie donnée, par un ensemble d'entités (protéines, ARN messager, métabolites, etc.) et d'activités (transcription, traduction, phosphorylation, etc.).

Par ailleurs, les explications et les prédictions fournies par la médecine des réseaux ne reposent pas que sur l'analyse des réseaux. On a ainsi vu que l'analyse du diseasome nécessite à plusieurs reprises d'importer plusieurs informations biologiques sur ce que font les gènes, la façon dont ils sont exprimés temporellement et spatialement, le type de produits ou d'interactions dans lesquels ils sont impliqués. L'hypothèse locale de la médecine des réseaux selon laquelle deux gènes qui sont impliqués dans la même maladie sont fonctionnellement reliés est d'ailleurs un exemple d'hypothèse mécaniste, tout comme l'hypothèse du module fonctionnel de la maladie qui suppose une série d'entités et d'activités qui permettent de comprendre l'ensemble des processus continus qui entrainent l'apparition de la maladie.

Nous ne pouvons pas nous prononcer ici sur la complémentarité des explications topologiques et mécanistes en général, comme l'ont fait récemment plusieurs philosophes (Brigandt, 2013 ; Levy et Bechtel, 2013), mais cette complémentarité est bien démontrée dans le cas de la médecine des réseaux et de sa théorie génétique.

6.3.4. En quoi la théorie génétique de la médecine des réseaux explique-t-elle la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité ?

La théorie génétique de la médecine des réseaux permet d'expliquer à la fois la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité. Ces trois points ayant déjà été abordés précédemment, nous ne faisons ici que récapituler notre propos.

(1) L'explication de la variabilité interindividuelle dépend de la notion de module fonctionnel de la maladie : en fonction de l'anomalie génétique principale, de l'ensemble des gènes modificateurs et du type de mutation (allèle non fonctionnel, allèle ayant acquis une autre fonction, allèle avec perte de fonction partielle, etc.), le module fonctionnel de la maladie peut être perturbé de différentes façons et différentes symptômes apparaîtront qui connaîtront une évolution différente dans le temps et qui bénéficieront de traitements différents, en fonction de l'interaction de ces déterminants génétiques avec les autres déterminants du module fonctionnel de la maladie (déterminants environnementaux, phénotypes intermédiaires, etc.).

(2) L'explication de la variabilité intra-individuelle est liée à la notion de phénotype moléculaire idiosyncrasique âge-dépendant. Pour une maladie chronique, par exemple, c'est cette notion de modules fonctionnels qui évoluent dans le temps en fonction de l'histoire personnelle d'un individu et du type de perturbations auxquelles il est exposé, qui est susceptible d'expliquer qu'une même maladie se présente différemment au cours de la vie d'un individu, que ce soit en termes de symptômes, de facteurs déclenchants, d'évolution ou de traitement.

(3) L'explication de la comorbidité est liée à la caractérisation du diseasome comme un réseau invariant et à l'hypothèse selon laquelle les modules fonctionnels d'une maladie peuvent se chevaucher. À partir du moment où deux maladies partagent dans leur physiopathologie un même gène, une mutation du gène peut avoir des conséquences sur les deux maladies et augmenter leur probabilité de survenir ensemble chez un patient donné.

6.4. Les maladies infectieuses du point de vue de la médecine des réseaux

Nous avons montré que la médecine des réseaux pouvait être comprise comme une théorie générale de la maladie, au sein de laquelle on pouvait identifier une *théorie génétique de la maladie* en général. Or, au chapitre 5, nous avons démontré qu'il existait une théorie génétique des maladies infectieuses qui pouvait servir de modèle à une *théorie génétique des maladies*. Nous nous trouvons donc confrontés à une difficulté : comment résoudre notre alternative élaborée dans la deuxième partie de ce travail, selon laquelle la généticisation reflète soit une théorie génétique des maladies, soit une théorie génétique de la maladie ? Comment comprendre cette coexistence ? Faut-il voir ces deux théories comme des théories complémentaires ou comme des théories concurrentes ?

Pour répondre à cette question, nous nous proposons de revenir sur l'exemple des maladies infectieuses – mais au prisme de la médecine des réseaux. Notre objectif est ici triple : il s'agit de (1) comparer les explications que théorie génétique des maladies infectieuses et médecine des réseaux donnent de la classe des maladies infectieuses, (2) comparer comment ces deux théories intègrent plus ou moins l'existence d'autres explications partielles (le microbiome, la génétique du parasite) des maladies infectieuses, (3) comparer comment ces deux théories traitent les cas-limites de maladies infectieuses (cancers viro-induits, maladies auto-immunes, etc.). Précisons cependant dès à présent qu'il n'y a pas à proprement parler de théorie génétique de la médecine des réseaux pour les maladies infectieuses : l'approche par la médecine des réseaux des maladies infectieuses est encore très récente.

6.4.1. L'approche systémique des infections virales

Étant donné le caractère très récent de l'application de la médecine des réseaux aux maladies infectieuses, nous proposons de concentrer notre analyse sur les travaux de l'équipe de Chantal Rabourdin-Combe et Vincent Lotteau à l'Université de Lyon Sud, qui ont développé une approche systémique des infections virales. Deux raisons principales motivent ce choix.

Tout d'abord cette équipe fait partie de l'EISBM (*European Institute for Systems Biology and Medicine*), qui a été fondé en 2012 à Lyon Sud par Charles Auffray et qui a

l'ambition d'être l'équivalent des deux centres internationaux de biologie des systèmes que sont Seattle (dirigé par le Pr Leroy Hood) et Shanghaï (dirigé par le Pr Zhu Chen). Ces trois auteurs sont activement impliqués dans le projet de médecine des 4P (prédictive, préventive, personnalisée et participative), que nous avons déjà évoquée au début de ce chapitre (Auffray *et al.,* 2009). Leurs travaux reposent d'ailleurs fortement sur la médecine des réseaux telle qu'elle a été développée par Albert-László Barabási (leurs travaux sont ainsi cités sur le site internet de Barabási). C'est donc une équipe d'excellence, internationalement reconnue, qui entretient des liens forts avec la biologie et la médecine des systèmes et la médecine des réseaux.

Par ailleurs, cette équipe développe depuis plusieurs années une approche de biologie et de médecine des systèmes appliquée à l'infectiologie et plus précisément aux infections virales. Parmi les différents pathogènes des maladies infectieuses, les virus ont l'avantage de présenter des génomes relativement compacts, comportant un petit nombre de gènes, ce qui explique que de nombreux génomes viraux ont été séquencés. A ce titre, cette équipe participe au projet I-MAP (*Infection-mapping project*) dont le but est de *« fournir une vision globale des infections virales au niveau des protéines en cartographiant les interactions d'un large nombre de protéines virales avec des protéines de l'hôte. La détection et la cartographique ont été pensées pour répondre à des questions précises, telles que la virulence / atténuation, la notion de barrière d'espèce, l'identification de cibles thérapeutiques, la chronicité et le risque de développer un cancer. »* (de Chassey *et al.,* 2008, p. 2). Dans ce contexte, cette équipe a contribué au développement deux bases de données en ligne VirOrféome (Pellet *et al., 2010)* et VirHostNet (Navratil *et al.,* 2009).

Un ORFéome correspond à l'ensemble des phases de lecture ouvertes potentielles d'un organisme. Une phase de lecture ouverte désigne une séquence linéaire de nucléotides de l'ARN messager, comprise entre un codon d'initiation de la traduction et un codon d'arrêt de la traduction des protéines. Une phase de lecture ouverte ne contient donc ni les promoteurs, ni les séquences 5' et 3' non codantes des ARN messagers, ni les introns qui sont les séquences épissées de l'ARN messager et qui sont absentes des transcrits matures. L'ORFéome répond à un problème posé par la protéomique classique, qui se focalise en général sur l'étude des protéines exprimées *in vivo,* mais qui peine parfois à détecter les protéines les plus faiblement exprimées

par un organisme. L'ORFéome est une technique de protéomique inverse : plutôt que de partir des protéines exprimées *in vivo*, on identifie les phases ouvertes de lecture potentielles d'un organisme, puis on introduit ces séquences dans un hôte cellulaire qui permet de les produire en grande quantité : l'ORFéome permet donc d'avoir un meilleur accès à l'ensemble des protéines produites par un organisme, quel que soit le niveau d'expression de ces protéines dans l'organisme. VirORFéome est donc un ensemble d'ORFéomes viraux qui comprend l'ensemble des phases de lecture ouvertes potentielles des virus. VirHostNet est une base de données qui répertorie l'ensemble des interactions protéines-protéines entre des protéines virales et d'autres protéines virales et entre des protéines virales et l'interactome humain (interactions protéines virus – protéines de l'hôte). Cette base de données a été reconstruite à partir de la base de données VirORFéome et de l'intégration de différentes bases de données publiques sur les interactions protéines-protéines virus-virus et les interactions protéines-protéines virus-hôte, toutes les interactions ayant été vérifiées manuellement. A son lancement, VirHostNet contenait 2 671 interactions protéines virales – protéines hôte, analysant plus de 180 espèces virales différentes et un des plus larges interactomes humains reconstruits (10 672 protéines et 68 252 interactions protéines-protéines humaines non redondantes).

Ces deux bases de données ont permis à cette équipe de développer une approche transversale de biologie des systèmes appliquée aux maladies infectieuses (voir Figure 58).

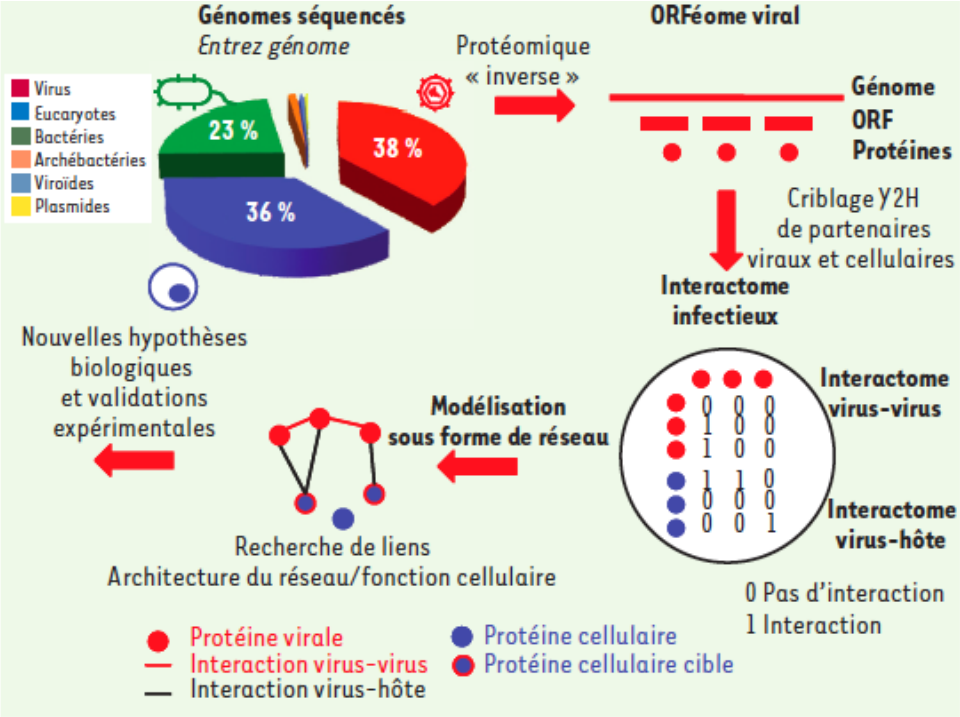


Figure 58 : Du génome au réseau d'interactions protéiques du virus, une approche transversale de biologie moléculaire systémique (Navratil *et al.*, 2010, p. 605). Une répartition taxonomique des génomes séquencés est tout d'abord donnée. Ces génomes sont utilisés comme un canevas (par une approche de protéomique inverse) pour l'annotation des protéines encodées et la genèse de l'ORFéome viral. L'ORFéome viral est utilisé pour le criblage systématique de l'interactome virus-hôte et hôte virus. Un exemple théorique d'un ensemble de recherche de 18 combinaisons concernant 3 protéines vitales (en rouge) et 3 protéines de l'hôte (en bleu) est présenté : on note 1 s'il y a interaction, 0 s'il n'y a pas d'interaction. Les interactions expérimentales (arêtes rouges) sont intégrées au sein d'une modélisation du réseau d'interactions entre les protéines du virus (nœuds rouges) et celles de l'hôte (nœuds bleus). L'étude systématique des réseaux d'interactions protéine-protéine permet l'identification de liens entre l'architecture du réseau et le fonctionnement cellulaire. Cette approche conduit à la proposition de nouvelles hypothèses biologiques testables expérimentalement.

L'objectif de cette approche systémique est de reconstruire l'infectome, l'ensemble des interactions moléculaires et fonctionnelles (par exemple les régulations génétiques, les interactions protéine-protéine, les réaction métaboliques) entre les composants des virus (génome, capside, enveloppe virale, etc.) et ceux de la cellule hôte (le cytosquelette, le noyau, les membranes, les organites, etc.). À partir de cet infectome, il est possible de reconstruire « un modèle de cellule virtuelle infectée » (voir Figure 59), c'est-à-dire d'établir l'ensemble des interactions entre les protéines virales et les protéines de l'hôte et d'appliquer les concepts usuels de la théorie des réseaux pour comprendre l'architecture de ces interactions. C'est ce que cette équipe a fait pour plusieurs virus et entre autres pour le virus de l'hépatite C (de Chassey *et al.*, 2008).

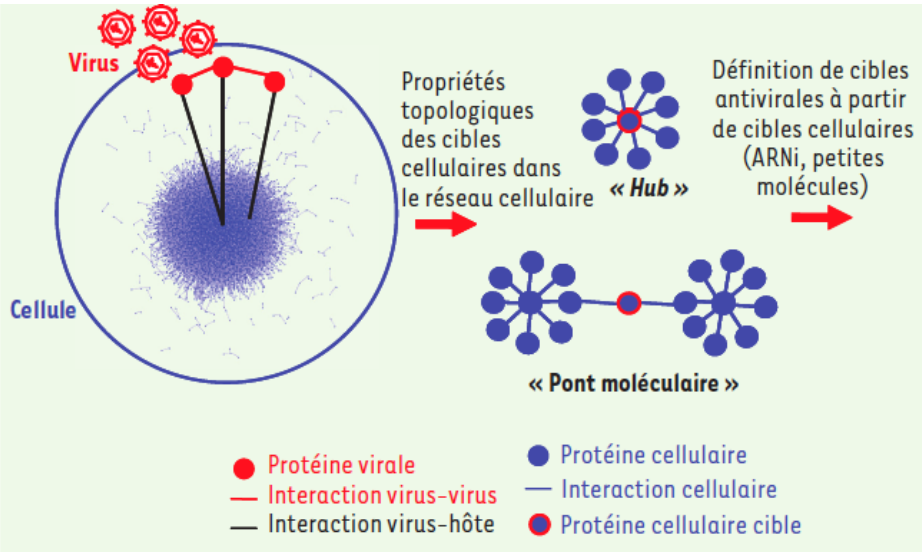


Figure 59 : La cellule infectée virtuelle, un modèle systémique de l'infection (Navratil *et al.*, 2010, p. 604). Les cibles cellulaires du virus sont projetées au sein du réseau d'interactions protéine-protéine de la cellule de l'hôte. Les interactions entre les protéines cellulaires sont représentées en bleu, les protéines virales sont représentées en rouge. Une analyse topologique de ce réseau permet d'assigner des propriétés topologiques particulières à certaines protéines du réseau. Les *hubs* sont des protéines cellulaires ciblées par le virus et qui ont une connectivité k très élevée dans le réseau cellulaire de l'hôte. Les « ponts moléculaires » sont des protéines cellulaires qui sont peu connectées mais qui relient des *hubs* entre elles et qui sont donc très centrales dans le réseau cellulaire de l'hôte. À partir de l'identification des propriétés topologiques des cibles cellulaires dans le réseau cellulaire, il est possible de définir des cibles antivirales.

Il est également possible de confronter l'infectome au diseasome et c'est ce que cette équipe a réalisé dans un article intitulé « When the human viral infectome and diseasome networks collide : towards a systems biology platform for the aetiology of human diseases » (Navratil *et al.,* 2011), et sur lequel nous nous appuierons tout particulièrement, pour comparer la façon dont les maladies infectieuses sont expliquées par la théorie génétique des maladies infectieuses et par la médecine des réseaux.

6.4.2. Quand l'infectome rencontre le diseasome

Étant donné la multiplicité et la complexité des méthodes mises en jeu dans cet article, nous nous contenterons d'expliciter l'objectif principal, la méthodologie générale et les conclusions auxquelles les auteurs parviennent. L'objectif principal de cet article est de répondre à deux grandes questions : quelles sont les protéines cellulaires ciblées par les infections virales et quelles sont les liens entre les protéines

ciblées par les infections virales et l'apparition de certaines maladies comme les cancers viro-induits, les maladies chroniques et les maladies auto-immunes ?

La première question concerne les cibles des infections virales. Les virus ont en effet en commun de ne pas posséder une machinerie cellulaire leur permettant de se répliquer. Lorsqu'ils infectent une cellule, ils ont donc besoin de détourner la machinerie cellulaire de l'hôte pour permettre la transcription et la traduction de leur génome et ainsi garantir leur réplication. Quelles sont donc les protéines de l'hôte que les virus attaquent en priorité et peut-on en déduire des cibles thérapeutiques antivirales spécifiques ? Pour répondre à cette question, les auteurs ont reconstruit à partir de la base de données VirHostNet un premier réseau, le réseau de l'infectome humain (HIN ou *Human Infectome Network*) qui répertorie l'ensemble des interactions physiques entre protéines virales et protéines de l'hôte.[60] Le réseau de l'infectome humain comporte d'une part 416 protéines virales venant de différentes espèces de virus (le protéome viral, représenté en rouge dans la Figure 60) et d'autre part, 20 219 protéines humaines (le protéome humain, représenté en bleu dans la Figure 60). Le protéome viral et le protéome humain sont reliés par 2 009 interactions protéines-protéines virus-hôte identifiées par des arêtes rouges dans la Figure 60. Toutes les protéines du protéome ne sont donc pas ciblées par les protéines virales : lorsqu'elles ne sont pas ciblées par des protéines virales, elles sont représentées par de simples carrés bleus, lorsqu'elles sont directement ciblées par des protéines virales, elles sont appelées protéines ciblées ou TP (*targeted proteins*) et lorsqu'elles sont reliées à une autre protéine qui est directement ciblée par une protéine virale, elles sont appelées protéines voisines d'une protéine ciblée ou TP-N (*targeted proteins-neighbour*).

La seconde question concerne les relations qu'entretient le réseau de l'infectome humain avec l'ensemble des protéines impliquées dans des maladies humaines et que les auteurs appellent ici diseasome[61]. Pour répondre à cette question, les auteurs ont combiné le premier réseau obtenu, le réseau de l'infectome humain au réseau du diseasome : comme on le voit dans la Figure 60, les maladies sont représentées par des nœuds noirs et sont reliées par des arêtes noires aux protéines

[60] Seules les protéines interconnectées (interactions protéines virus-virus, interactions protéines virus-hôte, interactions protéines hôte-hôte) sont représentées.

[61] Ce terme n'est donc pas employé dans le même sens que dans les articles que nous avons étudiés au chapitre 6, où le diseasome représentait l'ensemble des gènes impliqués dans les maladies humaines.

du protéome humain qui sont associées à ces maladies, d'après les associations protéines-maladies, récupérées à partir de la base de données OMIM. Les protéines du protéome humain qui sont associées à des maladies sont renommées DRP (*disease-related proteins*). Parmi ces DRP, certaines sont la cible directe des virus (auquel cas, elles sont entourées d'un carré rouge) et d'autres sont connectées à une autre protéine humaine, qui est elle-même la cible directe d'un virus (auquel cas, elles sont dites indirectement ciblées par les virus et sont entourées d'un carré orange). Par ailleurs, les auteurs ont ensuite étudié dans le détail des sous-parties de ce réseau de l'infectome et du diseasome humain (HIDN ou *Human Infectome-Diseasome Network*) pour répondre à des questions plus précises, concernant les relations entre le réseau de l'infectome humain et les maladies chroniques, les cancers viro-induits et les maladies auto-immunes.

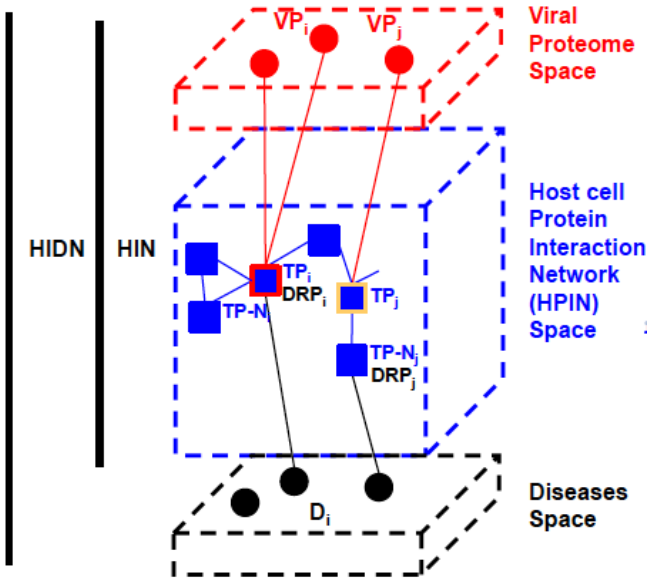


Figure 60 : Le réseau de l'infectome humain et le réseau de l'infectome-diseasome humain (Navratil *et al.*, 2011, p. 3). Le réseau de l'infectome humain (HIN) représente l'ensemble des interactions possibles entre des protéines virales (en rouge) et des protéines de l'hôte (en bleu), représentées par des arêtes rouges, ainsi que l'ensemble des interactions protéines-protéines de l'hôte (arêtes bleues). Les protéines de l'hôte directement connectées à une protéine virale sont entourées d'un carré rouge et sont appelées TP (*targeted protein*). Les protéines de l'hôte qui sont connectées à une protéine virale *via* une interaction protéine – protéine de l'hôte sont entourées d'un carré orange et sont appelées TP-N (*targeted protein – neighbour*). Le réseau de l'infectome-diseasome humain (HIDN) est le réseau des interactions entre les maladies (nœuds noirs) et les protéines de l'hôte ciblées par les protéines virales. Lorsqu'une maladie (en noir) est associée directement à une protéine TP, celle-ci est appelée DRP (*disease related protein*). Lorsqu'une maladie est associée à une protéine TP, *via* une interaction protéine-protéine de l'hôte, la protéine en question est appelée DRP-N (d*isease-related protein neighbour*).

6.4.2.1. L'analyse du réseau de l'infectome humain

Quelles protéines de l'hôte les virus ciblent-ils ? Une première analyse topologique générale du réseau de l'infectome humain (Figure 61) révèle que les 416 protéines virales ont pour cible 5% des protéines du protéome humain, c'est-à-dire 1 148 protéines ciblées. 32% de ces protéines cibles de l'hôte interagissent avec au moins une protéine virale. Une seconde analyse topologique se concentre autour du paramètre de connectivité (c'est-à-dire le nombre de degrés d'un nœud donné) et autour du paramètre de centralité (nombre de chemins le plus courts qui passent par une protéine), qui permet d'estimer le flux théorique d'information potentiellement contrôlé par une protéine donnée. Ces analyses topologiques permettent de conclure que les protéines virales interagissent préférentiellement avec des protéines cibles de l'hôte qui sont hautement connectées (connectivité des protéines cibles de l'hôte = 38 contre connectivité des protéines non cibles de l'hôte = 10), centrales et très interconnectées entre elles, c'est-à-dire avec des protéines *hubs* du réseau cellulaire de l'hôte. Le vaste spectre des protéines de l'hôte ciblées par les virus (que reflète la connectivité élevée des protéines virales) est assez cohérent avec la pléiotropie fonctionnelle des protéines virales et avec la diversité des stratégies moléculaires que les virus ont acquis au cours de l'évolution pour s'emparer de la machinerie cellulaire de l'hôte et échapper aux défenses antivirales de l'hôte.

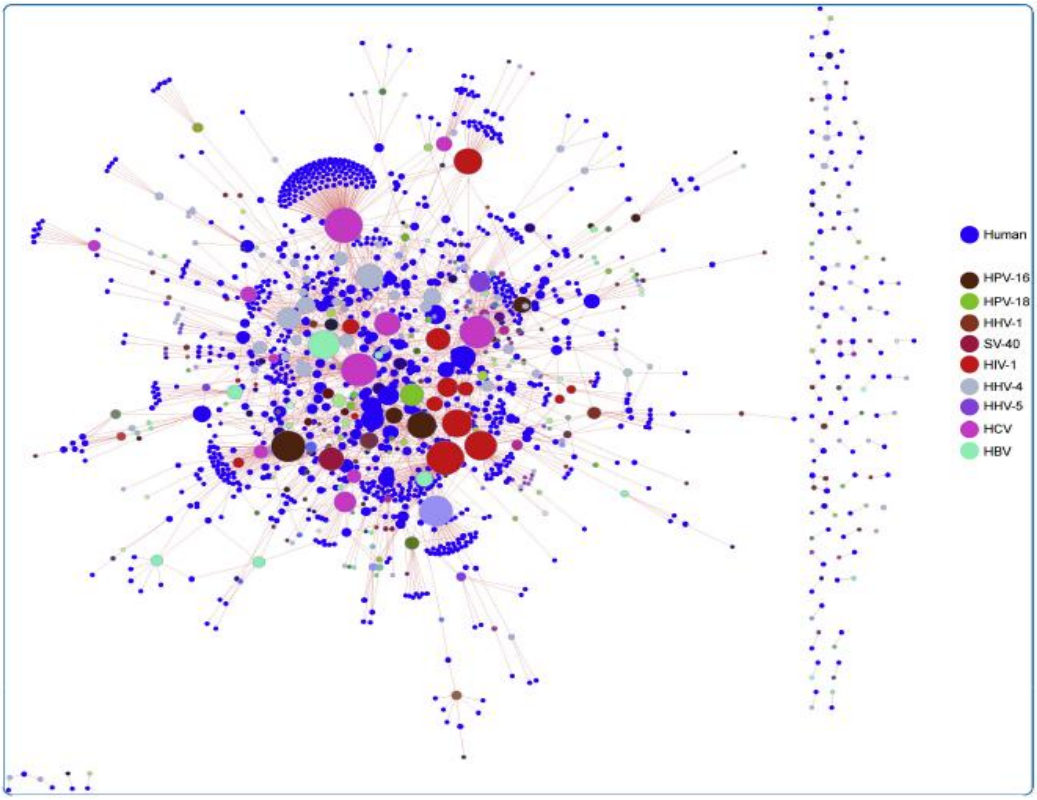


Figure 61 : Représentation du réseau de l'infectome humain (Navratil *et al., 2011,* p. 1). Les nœuds bleus représentent les protéines cellulaires humaines. Les autres nœuds colorées représentent les protéines virales des principales espèces virales impliquées et parmi lesquelles on retrouve la famille des papillomavirus (HPV-16 et HPV-18, le virus de l'hépatite B (HCV), le virus de l'hépatite C (HCV), etc. Deux nœuds sont connectés entre eux si les protéines ont une interaction physique.

Les auteurs ont ensuite cherché à déterminer si les protéines virales ciblent des modules particuliers du réseau de protéines de l'hôte. Pour ce faire, ils ont décomposé le protéome humain en sous-groupes de modules topologiques en utilisant un algorithme qui permet d'isoler des modules cohérents. Cette décomposition aboutit à 311 modules interconnectés comprenant 6 152 protéines. 38% de ces modules (c'est-à-dire 118 modules) sont ciblés par au moins une protéine virale. 17 modules (représentés en rouge dans la Figure 62) sont particulièrement ciblés (associés à des fonctions comme le transport de protéines, l'attachement du virion aux récepteurs de surface des cellules de l'hôte, différenciation cellulaire, régulation de l'apoptose, etc.). Les autres modules (noirs) ne sont pas ou très peu ciblés. 541 protéines appartiennent à plusieurs modules et jouent le rôle de protéines de liaison entre les modules (*module linkers*). 20% de ces protéines de liaison sont au moins ciblées par une protéine virale,

ce qui suggère que les virus ont tendance à perturber des protéines qui jouent le rôle de carrefour entre plusieurs modules fonctionnels.

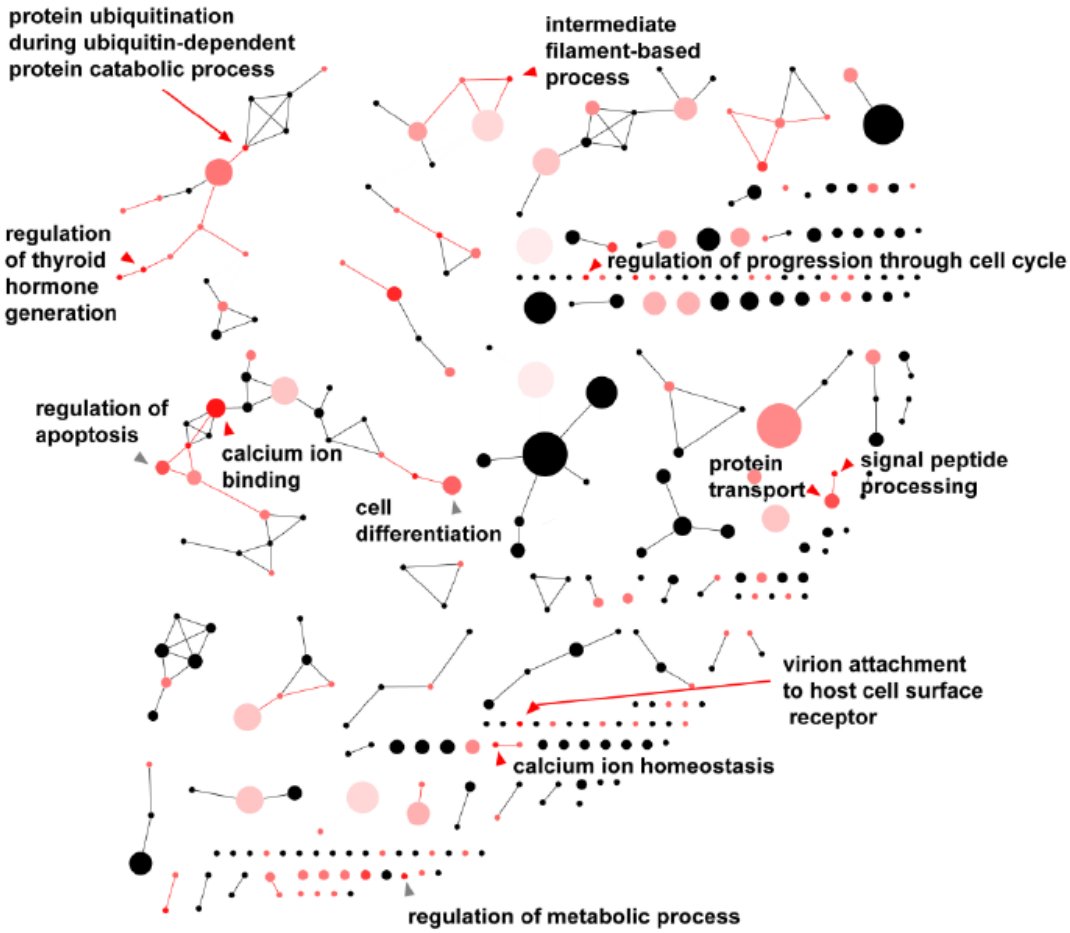

protein ubiquitination during ubiquitin-dependent protein catabolic process
intermediate filament-based process
regulation of thyroid hormone generation
regulation of progression through cell cycle
regulation of apoptosis
calcium ion binding
cell differentiation
protein transport
signal peptide processing
virion attachment to host cell surface receptor
calcium ion homeostasis
regulation of metabolic process

Figure 62 : Décomposition modulaire du réseau de l'infectome humain (Navratil *et al.*, 2011, p. 5). La décomposition de l'HIN par l'algorithme CFinder permet d'identifier plusieurs modules interconnectés de protéines (nœuds) et de protéines de liaisons entre les modules. Les modules de protéines et les protéines de liaisons entre les modules sont colorés en fonction de l'intensité de l'attaque virale. Les modules les plus attaqués sont représentés en rouge, les modules les moins ciblés sont représentés en noir. Les fonctions biologiques des protéines les plus connectées sont notées et fléchées.

Les premières analyses topologiques semblaient montrer que les protéines virales étaient connectées en priorité à des protéines très connectées de l'hôte (*des protéines hubs*). Mais le protéome humain est de façon général assez peu connecté : 50% des protéines humaines du protéome humain n'interagissent qu'avec un seul partenaire cellulaire. C'est pourquoi les auteurs ont essayé de caractériser le type de protéines peu connectées que les protéines virales ciblent (en dehors des protéines très connectées du réseau de l'hôte qui représentent la majorité de leurs cibles). Parmi

les protéines de l'hôte faiblement connectés, les auteurs ont observé que les protéines virales ciblent en particulier des protéines appelées « ponts moléculaires » (*bridging proteins*) ou « goulots d'étranglement » (*bottleneck proteins*). Ces protéines, représentées dans la Figure 59, ont peu de partenaires moléculaires mais sont aussi caractérisées par une centralité élevée par rapport aux autres protéines du protéome. Autrement dit, ces protéines sont peu connectées au reste de l'interactome mais elles permettent de relier des groupes de protéines très connectées au reste du réseau, c'est-à-dire de relier des *hubs* entre eux : *« Nous avons fait l'hypothèse que cette catégorie de protéines qui fait des ponts entre des parties hautement connectées des réseaux cellulaires avec un nombre limité d'événements de liaisons pourrait être des points de contrôle critiques pour le contrôle et la régulation du flux d'information cellulaire »* (Navratil *et al.,* 2011, p. 6).

En conclusion, l'analyse du réseau de l'infectome humain permet de distinguer deux types d'attaques virales du protéome humain : pour prendre le contrôle de la machinerie cellulaire de leur hôte, les protéines virales peuvent soit cibler directement les modules fonctionnels centraux, soit contrôler comment ces fonctions cellulaires sont co-régulées en ciblant les protéines qui jouent un rôle de ponts moléculaires, ce que les auteurs appellent une « attaque furtive » (*stealth-attack*).

6.4.2.2. L'analyse du réseau de l'infectome-diseasome humain

Passons à présent à l'analyse des rapports entre le réseau de l'infectome humain et le diseasome humain : 1 729 DRP ou protéines associées à 1 908 maladies répertoriées dans OMIM ont été intégrées au réseau de l'infectome humain. Une première analyse montre que 12% des 1 148 protéines de l'hôte ciblées par des protéines virales sont associées à au moins une maladie : une protéine humaine qui interagit avec une protéine virale a donc environ deux fois plus de chances d'être impliquée dans une maladie humaine qu'une protéine humaine qui n'est pas ciblée par un virus. Ce pourcentage passe à 75% quand on prend en compte les protéines voisines des protéines ciblées de l'hôte. Inversement, environ 50% des 1 908 maladies répertoriées dans OMIM sont connectées soit directement soit indirectement à au moins un virus. 34 maladies en particulier ont été identifiées comme significativement

liées à des virus, la majorité de ces maladies étant des cancers (cancer du poumon, carcinome hépatocellulaire, cancer du sein, leucémie…) et des maladies neurodégénératives (maladie de Parkinson, etc).

À partir de ces données, les auteurs ont construit le réseau de l'infectome-diseasome humain (HIDN) comme un graphe bipartite composé de deux types de nœuds : les nœuds noirs sont les maladies et les nœuds colorés sont les virus (voir Figure 63). Dans l'HIDN, une maladie et un virus sont connectés par une arête si au moins une protéine reliée à une maladie donnée est ciblée par au moins une protéine encodée par un virus. Une version étendue de ce réseau HIDN a ensuite été proposée qui inclut les voisins directs des protéines reliées à la maladie qui sont ciblées par un virus.

Figure 63 : Le réseau infectome-diseasome humain (version non étendue) (Navratil *et al.,* 2011, p. 9). Les maladies sont représentées par des nœuds noirs, les virus par des nœuds rouges.

Dans l'HIDN, les cancers occupent une position centrale avec une prédominance des cancers liées aux virus à ADN, comme le cancer du col et les papillomavirus. Une exception notable est le cas du virus de l'hépatite C qui est relié essentiellement à des maladies chroniques comme la cirrhose, le carcinome

hépatocellulaire, qui sont des complications bien connues de l'hépatite C. La version étendue de l'HIDN a été appliquée en particulier au réseau des protéines du virus de l'hépatite C : pour prendre en compte les interactions potentielles entre le protéome du virus de l'hépatite C, les protéines de l'hôte voisines des protéines ciblées directement par le virus ont été prises en compte. En utilisant différents algorithmes, les auteurs ont identifié les protéines directement ciblées et les protéines voisines de l'hôte associées à chacune de ces complications chroniques, confirmant leur analyse précédente du réseau des protéines du virus de l'hépatite C (de Chassey *et al., 2008)*.

Une des analyses les plus intéressantes du réseau de l'infectome-diseasome humain concerne les relations entre infections virales et maladies auto-immunes. De nombreuses maladies auto-immunes sont suspectées d'avoir une origine virale, mais celle-ci n'est pas toujours facile à mettre en évidence. À partir du réseau de l'infectome-diseasome humain, les auteurs ont donc reconstruit une version étendue de l'HIDN (comprenant donc les protéines voisines des protéines de l'hôte directement reliées à la maladie) spécifique des maladies auto-immunes répertoriées dans OMIM (Figure 64).

Dans ce réseau de l'infectome-diseasome auto-immun, le diabète de type 1, maladie auto-immune qui implique la destruction des cellules béta du pancréas (qui produisent l'insuline) par des anticorps auto-immuns et dont l'origine virale fait l'objet de nombreux débats, a une position centrale. La version étendue de l'HIDN permet de mettre en évidence une connexion directe entre la famille des Paramyxovirus et la diabète de type 1 via une interaction entre la protéine virale V et la protéine de l'hôte IFIH1, qui a été récemment identifiée par deux associations pangénomiques indépendantes comme un locus de susceptibilité du diabète de type 1. Par ailleurs, les auteurs identifient 38 protéines de l'hôte qui sont ciblées par des virus et qui sont directement connectées au diabète de type 1 (que les auteurs appellent des protéines « prises la main dans le sac »). Les gènes codant ces 38 protéines pourraient donc être de nouveaux gènes de susceptibilité au diabète de type 1. Ces protéines ont toutes des fonctions fortement associées au système immunitaire, elles sont par ailleurs toutes exprimées par les cellules beta et 50% de ces protéines sont produites par des gènes localisés dans des régions précédemment identifiées comme des régions de susceptibilité génétique au diabète de type 1 chez l'homme et la souris.

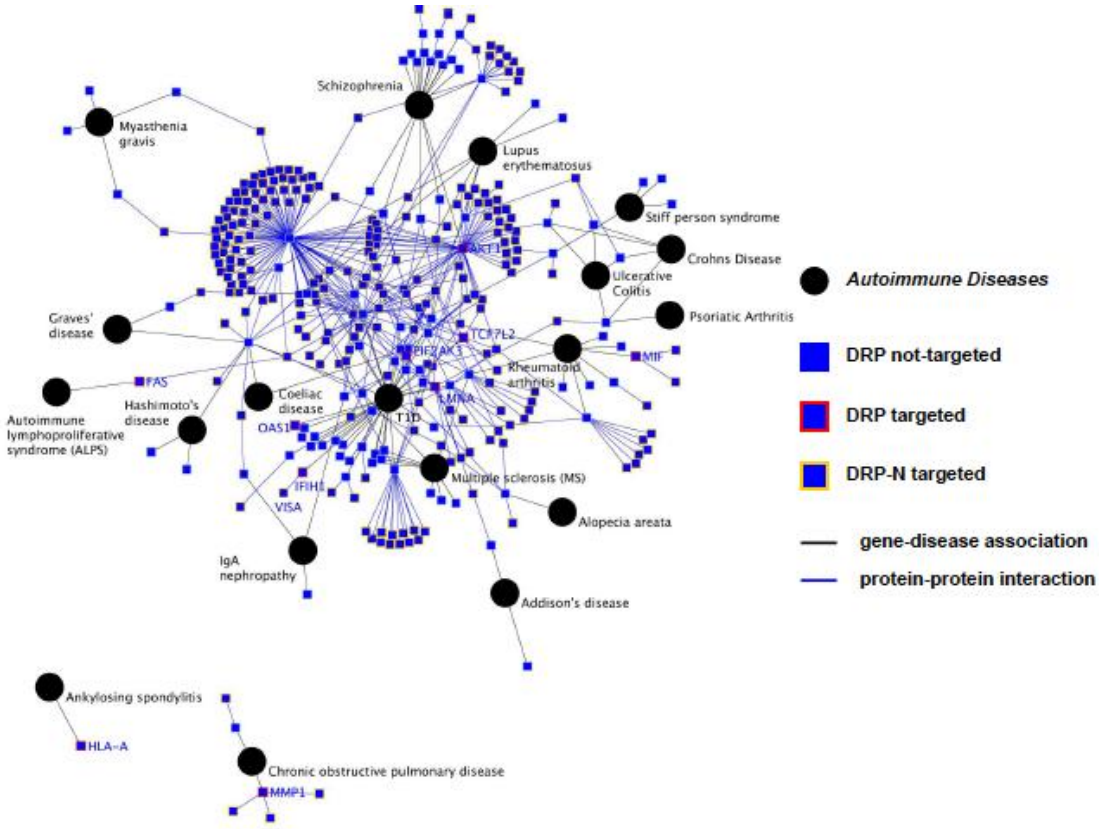


Figure 64 : Le réseau de l'infectome et du diseasome auto-immun (Navratil *et al.,* 2011, p. 11). Deux types de nœuds sont représentées : les maladies sont des cercles noires et les protéines cellulaires de l'hôte sont des carrés bleus. Les protéines cellulaires de l'hôte peuvent être directement connectées à la maladie (DRP) ou indirectement connectées à travers une protéine voisine (DRP-N). Les interactions protéines-protéines entre les protéines cellulaires de l'hôte sont représentées par des arêtes bleues. Les associations gènes-maladies sont représentées par des arêtes noires. Les protéines associées à des maladies (DRP) qui sont ciblées par des virus sont entourées d'un carré rouge (par exemple, IFIH1). Les DRP-N ciblées par des virus sont entourées d'un carré orange (par exemple, VISA).

6.4.2.3. Conclusions de l'analyse du réseau de l'infectome humain et du réseau de l'infectome-diseasome humain

Que nous apprend cet article ? Les conclusions des auteurs se concentrent surtout sur les applications thérapeutiques possibles de leurs résultats. Les protéines centrales très connectées (protéines *hubs*) qui sont attaquées directement par les virus pourraient sembler des cibles thérapeutiques idéales, mais lorsque les auteurs réalisent des perturbations *in silico* de ces protéines, l'ensemble du protéome humain est très rapidement déstabilisé, précisément parce que ces protéines *hubs* ont des fonctions centrales dans la machinerie cellulaire de l'hôte. Des antiviraux qui cibleraient les protéines *hubs* de l'hôte risqueraient donc d'être très toxiques pour

l'hôte. En revanche, les protéines qui jouent le rôle de pont moléculaire, les *bridging proteins*, pourraient être des cibles thérapeutiques beaucoup plus adaptées pour au moins trois raisons. Tout d'abord, elles sont directement et massivement ciblées par les protéines virales. Ensuite, leur importance pour le réseau du protéome humain est moindre (le protéome humain est plus robuste contre la délétion de protéines-ponts moléculaires que contre la délétion de protéines *hubs*), ce qui veut dire que des molécules antivirales ciblant ces protéines seraient mieux tolérées par l'hôte. Enfin, l'un des problèmes majeurs des thérapies antivirales sont les capacités d'adaptation évolutive des virus : ce type d'adaptation évolutive est peu probable dans le cas des protéines de type ponts moléculaires.

6.4.3. Les maladies infectieuses au sein de la TGMI et au sein de la médecine des réseaux

Que nous dit cet article d'une éventuelle théorie génétique de la médecine des réseaux appliquée aux virus ? L'approche systémique des infections virales est encore très récente et principalement spéculative. L'apport essentiel de cet article est probablement de suggérer différentes façons d'explorer et d'utiliser la base de données VirHostNet qui, bien qu'elle soit biaisée en faveur des virus les mieux étudiés, est une base de données qui répertorie manuellement, avec un haut degré de confiance et aussi exhaustivement que possible, les interactions entre les protéines virales et les protéines de l'hôte. Le fait de comparer l'infectome au diseasome et d'incorporer non seulement les interactions directes, mais aussi les interactions indirectes entre des protéines de l'hôte associées à des maladies et des protéines virales, a un pouvoir heuristique potentiel fort, comme le suggèrent les 38 nouvelles protéines candidates mises en évidence dans l'apparition du diabète de type 1. Ce pouvoir heuristique sera d'autant plus fort que le diseasome a été en l'occurrence reconstruit à partir d'OMIM, dont on a souligné qu'elle répertoriait encore majoritairement des gènes et des protéines associés aux maladies monogéniques. En dépit de ces premiers résultats stimulants, il n'y a certainement pas pour l'instant de théorie génétique de la médecine des réseaux appliquée aux virus et cette approche n'est qu'un premier pas dans l'exploration des réseaux de protéines virales et des

relations entre les infections virales et d'autres types de maladies comme les cancers, les maladies auto-immunes, ou les maladies chroniques.

Cependant, et nous soulignons ici que notre position est tout à fait spéculative, il nous semble que plusieurs avantages très prometteurs de l'approche des maladies virales par la médecine des réseaux méritent d'être soulignés. La théorie génétique des maladies infectieuses proposée par Abel et Casanova et construite progressivement ces trente dernières années, présente certaines limites : elle ne précise pas le rôle de la génétique du parasite ou du microbiome et n'offre pas de cadre explicatif pour explorer ce que nous avons appelé les « cas-limites » de maladies infectieuses, comme les cancers viro-induits, les relations entre maladies infectieuses et maladies chroniques, les relations entre maladies infectieuses et maladies auto-immunes. Au contraire, l'approche des maladies virales, développée beaucoup plus récemment, même si elle est encore balbutiante, semble en mesure de développer les outils pour répondre à ces questions. C'est ainsi que Vincent Navratil, l'un des membres de l'équipe de Chantal Rabourdin-Combe et de Vincent Lotteau écrit :

> « Ainsi, dans un futur proche, l'intégration multi-échelles des données d' « -omiques » fusionnant des approches de biologie moléculaire systémique « ascendante » et « descendante » devrait permettre une modélisation fine des différents modes d'organisation et de réponse fonctionnelle de la cellule au cours de l'infection. Dans ce contexte, le modèle de la cellule infectée virtuelle pourra alors être décliné selon différentes pathologies, différents virus, différentes familles de protéines, etc. Finalement, l'étude systématique panpathogène et comparative inter-souches (souches virulentes / non virulentes, résistantes / non résistantes au traitement, etc.) et des empreintes et profils moléculaires devrait conduire au développement d'outils diagnostiques, notamment pour les maladies dont l'étiologie virale est incertaine (cancers, maladies auto-immunes, etc.) et celles qui sont associées à la chronicité d'une infection (le syndrome métabolique, par exemple). » (Navratil *et al., 2010,* p. 607–608)

Ainsi, la médecine des réseaux prend mieux en compte le rôle de la génétique du parasite puisqu'elle étudie directement les interactions entre les protéines virales et les protéines de l'hôte. L'article que nous avons présenté étudie les relations

globales de tous les virus connus et de toutes les protéines de l'hôte, mais on pourrait tout à fait imaginer, comme le fait Navratil, que le même type d'étude soit restreint à une seule famille de virus et à leurs différentes souches virulentes / non virulentes, sensibles ou résistantes au traitement, etc.

Par ailleurs, en 2012, les représentants du Projet Microbiome Humain estimaient avoir identifié près de 90% des différentes espèces de micro-organismes impliquées dans le microbiome. Une approche des maladies infectieuses par la médecine des réseaux pourrait tout à fait intégrer ces données pour comprendre le rôle du microbiome dans les maladies infectieuses et le lien entre rôle du microbiome et les « cas limites » de maladies infectieuses.

Enfin, les « cas limites » de maladies infectieuses sont pris en compte par cette approche, même si très peu de conclusions robustes ont été proposées jusqu'ici. Une meilleure exploration de ces questions pourrait amener à reclasser différemment certaines maladies cancéreuses, certaines maladies chroniques (comme la maladie de Parkinson, dont on a vu qu'elle était fortement associée aux infections virales) ou certaines maladies auto-immunes comme le diabète de type 1 comprises alors comme des maladies infectieuses, cette notion même devant être alors redéfinie.

L'approche des maladies infectieuses par la médecine des réseaux ne saurait cependant suffire à expliquer totalement l'ensemble des maladies infectieuses. Comme nous l'avons souligné à plusieurs reprises, la théorie génétique des maladies infectieuses est une explication mécaniste des maladies infectieuses, centrée sur la génétique de l'hôte. L'approche des maladies infectieuses par la médecine des réseaux est au contraire une explication topologique des maladies infectieuses, qui appelle d'autres explications mécanistes. C'est ainsi que Vincent Navratil insiste sur la nécessité de combiner leur approche avec d'autres techniques plus traditionnelles de biologie moléculaire pour parvenir à une explication mécaniste plus fine :

> « Comme tout modèle, cette description holistique du système infectieux n'est qu'un reflet très partiel de la réalité et de la complexité biologiques. Elle ne pourra en aucun cas se substituer aux modèles réduits classiquement développés en virologie. Ces deux approches, l'une réductionniste (l'étude des propriétés des composants), l'autre holistique (l'étude des propriétés émergentes du système d'interaction des composants) devront être utilisées

> de manière conjointe pour sans cesse s'enrichir mutuellement. Le développement synergique d'approches interdisciplinaires et transversales est l'une des clés de la réussite, comme l'atteste le nombre croissant d'instituts de recherche couplant les aspects de biologie systémique et de biologie analytique. » (Navratil *et al.,* 2010, p. 608)

Il n'y a pas d'opposition de principe entre la théorie génétique des maladies infectieuses développée par Casanova et Abel et les développements potentiels de l'approche systémique des maladies infectieuses : au contraire, ces deux explications sont complémentaires et répondent à des questions différentes avec des méthodes et des formes d'explication différentes. Ainsi, si cette triple comparaison semble suggérer que la médecine des réseaux permet d'expliquer les maladies infectieuses de façon plus globale (c'est-à-dire en prenant en compte un ensemble plus large de facteurs causaux impliqués dans les maladies infectieuses) et plus intégrée (c'est-à-dire en permettant de concilier plus facilement les différentes explications des maladies infectieuses) que ne le faisait la théorie génétique des maladies infectieuses, il semble néanmoins que la coexistence de ces deux types d'explication soit nécessaire pour rendre compte de la complexité des maladies infectieuses.

6.5. Conclusion du chapitre : limites méthodologiques et directions futures de la théorie génétique de la médecine des réseaux

Nous avons démontré que la médecine des réseaux était une théorie générale de la maladie, c'est-à-dire constituait une explication interactionniste co-constructionniste de la maladie en général, en se fondant sur une définition modulaire de la maladie. Nous avons également montré qu'au sein de cette théorie générale, il était possible de trouver une théorie génétique de la maladie qui explique l'origine génétique commune des maladies, renouvelle la distinction entre maladies monogéniques et maladies polygéniques et permet de prédire l'existence de nouveaux gènes de la maladie. Cette théorie génétique de la maladie mobilise des explications topologiques, sous-tendues ou complétées par des explications mécanistes. Enfin,

cette théorie génétique de la maladie permet de rendre compte de l'explication de la variabilité interindividuelle, de la variabilité intra-individuelle et de la comorbidité.

La médecine des réseaux apporte donc un véritable gain dans notre compréhension de la maladie en général et de la génétique des maladies en particulier. Elle n'en demeure pas moins une approche très récente de l'explication des maladies. Elle se heurte en particulier à certaines limites méthodologiques.

Nous avons déjà souligné ces limites méthodologiques : la médecine des réseaux repose dans sa démarche même sur la comparaison de données hétérogènes, incomplètes, collectées à partir de méthodes expérimentales différentes (modèles animaux, données cliniques, etc.), mobilisant des techniques différentes (modélisation *in silico*, validation biologique, associations pangénomiques, etc.) et à partir de bases de données qui ne partagent pas forcément le même système d'annotation. Cette démarche multiplie donc les risques de biais, même si ces biais tendront probablement à se réduire dans le futur, étant donné les efforts d'harmonisation considérables qui ont été entrepris dans cette époque de déluge de données.

Par ailleurs, la médecine des réseaux est une approche essentiellement *top-down*, dont on a déjà souligné (et Barabási le reconnaît lui-même) qu'elle nécessite pour être instanciée plus finement à l'échelle de chaque maladie qu'on poursuive des approches *bottom down*, telles celles développées par Alon Uri dans ses études des motifs récurrents présents dans les circuits biologiques.

Nous voudrions également souligner quelques ambiguïtés conceptuelles qui nous semblent mériter un travail de définition plus fin. Le concept de modularité, dont l'usage récurrent en biologie a été analysée en détail ailleurs (Mitchell, 2006), est empreint d'une ambiguïté fondamentale soulignée par Barabási lui-même : tantôt module topologique, tantôt module fonctionnel, tantôt module de la maladie. Le concept de fonction est également porteur d'ambiguïté : il semble la plupart du temps être employé dans le sens causal-mécaniste du terme, mais porte parfois son sens évolutionnaire. Un dernier concept nous paraît particulièrement crucial et c'est le terme de « gènes de la maladie humaine ». Ce concept est profondément instrumental, au sens où il a été forgé dans le cadre de l'analyse des réseaux de gènes impliqués dans les maladies humaines, pour répondre au problème de la multiplication des associations gènes-maladies. Mais ce concept instrumental semble acquérir de

plus en plus un contenu biologique, avec la mise en évidence de propriétés biologiques spécifiques aux gènes de la maladie humaine. Une question intéressante à explorer dans les développements de la théorie génétique de la médecine des réseaux serait de savoir à quel point ce concept de gène peut être un concept spécifique à la médecine et quelles seraient alors ses relations avec les concepts biologiques du gène classique, du gène moléculaire, et du gène de l'ère post-génomique.

Enfin, nous avons comparé l'explication des maladies infectieuses du point de vue de la théorie génétique des maladies infectieuses et du point de vue de la médecine des réseaux. S'il n'existe pas pour l'instant de théorie génétique des maladies infectieuses de la médecine des réseaux, nous pensons qu'une telle théorie pourrait intégrer le rôle de la génétique du pathogène et le rôle du microbiome dans l'explication des maladies infectieuses et rendre compte des cas-limites (cancers viro-induits, maladies chroniques, maladies auto-immunes liées à des infections), de façon plus globale et plus intégrée que ne le fait la théorie génétique des maladies infectieuses. Nous avons cependant insisté sur le fait que l'explication des maladies infectieuses ne permet pas de rendre compte des mêmes cibles explicatives que la théorie génétique des maladies infectieuses. La théorie génétique des maladies infectieuses est en outre beaucoup plus précise dans la description du rôle de la génétique de l'hôte que ne l'est l'approche des maladies infectieuses par la médecine des réseaux. Les théories génétiques régionales, comme la théorie génétique des maladies infectieuses, sont donc tout aussi nécessaires qu'une théorie génétique générale, comme celle qui se développe à l'heure actuelle dans la médecine des réseaux : dans la littérature biomédicale contemporaine, la coexistence de théories génétiques variées est nécessaire pour rendre compte à différents niveaux d'explication du rôle complexe que jouent les gènes dans toutes les maladies, et c'est ce point que nous allons développer dans notre conclusion générale.

Conclusion

Notre thèse principale est que la résolution du paradoxe de la génétique médicale contemporaine nécessite l'élaboration de théories génétiques. Mais la dernière partie de ce travail fait émerger une seconde thèse : la coexistence d'une pluralité de théories génétiques est nécessaire pour rendre compte des différents aspects de l'explication des maladies. Ce point nous paraît d'autant plus crucial que les développements récents de trois domaines de recherche, portant respectivement sur la fonction du génome non codant, le microbiome et l'épigénétique, semblent susceptibles de renouveler en profondeur dans les prochaines années notre compréhension de la génétique des maladies. Enfin, nous voulons proposer une réflexion plus générale sur le statut épistémologique de notre travail et plus précisément sur la façon dont la philosophie de la médecine gagnerait à s'emparer du concept de théorie médicale.

1. La coexistence d'une pluralité de théories génétiques est nécessaire pour résoudre le paradoxe de la génétique médicale contemporaine

1.1. L'approche du paradoxe par la sélection causale est une impasse

Au début des années 1960, le concept de maladie génétique correspond à celui de maladie monogénique : une mutation héréditaire mendélienne d'un gène cause une protéine mutée qui cause un phénotype pathologique. Depuis, un double mouvement s'est produit. D'une part, le concept de maladie monogénique a perdu sa simplicité apparente avec la découverte du rôle des gènes modificateurs, de l'hétérogénéité génétique et de l'hétérogénéité allélique, au point que d'aucuns clament que les maladies monogéniques n'existent pas ou qu'elles se réduisent à quelques exceptions. D'autre part, le concept de maladie génétique s'est progressivement étendu à des maladies non mendéliennes, non monogéniques, non

héréditaires, communes et complexes. Un grand nombre de variants alléliques impliqués dans des maladies complexes communes ont été identifiés. Progressivement, toute maladie associée à un variant allélique a été considérée comme génétique, quelle que soit l'origine (hérédité mendélienne, non mendélienne, mutation *de novo*) et quel que soit le type (anomalie chromosomique, mutation mitochondriale, variant rare, variant commun) de variation allélique impliquée. Cette extension du concept de maladie génétique a conduit à une généticisation progressive de toutes les maladies. Ce double mouvement de bouleversement du concept de maladie monogénique et de généticisation des maladies a abouti à ce que nous avons appelé « le paradoxe de la génétique contemporaine médicale » : toute maladie est considérée comme génétique alors même qu'il n'y a pas de consensus sur la définition du concept de maladie génétique.

Les philosophes ont majoritairement adopté la même stratégie pour résoudre ce paradoxe de la génétique médicale : ils réfutent la généticisation comme une extension génocentriste abusive du concept de maladie génétique et cherchent à redéfinir le concept de maladie génétique de façon plus stricte afin de pouvoir distinguer les maladies génétiques au sens strict (dans lesquelles le gène est la cause la plus importante) des autres maladies, considérées comme non génétiques. Différentes approches (l'approche cause-condition, l'approche différentialiste, l'approche pragmatique) que nous avons explorées dans le chapitre 2, sont alors utilisées pour justifier cette sélection causale du gène dans l'explication de certaines maladies.

Dans le contexte scientifique actuel que nous avons décrit, une telle stratégie pour résoudre le paradoxe de la génétique médicale est inadaptée à plusieurs titres. (1) Tout d'abord, aucune des approches mentionnées ne permet de justifier de façon satisfaisante la sélection causale du gène dans l'explication de certaines maladies et de parvenir à un concept cohérent de maladie génétique. (2) Aborder le paradoxe de la génétique médicale sous l'angle de la sélection causale revient d'ailleurs précisément à défendre, au moins pour certaines maladies considérées comme des maladies génétiques au sens strict, la position génocentriste que la résolution du paradoxe prétend pourtant réfuter. (3) Une telle stratégie ne permet pas de rendre compte de façon unifiée du rôle des gènes dans l'ensemble des maladies. Elle ne dit rien par exemple du rôle des gènes dans les maladies considérées comme non génétiques,

c'est-à-dire les maladies communes complexes. (4) Une telle stratégie simplifie à l'extrême la matrice causale de l'explication des maladies en se fondant sur une dichotomie gène / environnement qui est inadaptée pour décrire la complexité de la relation entre le génome codant et le phénotype et la diversité des facteurs causaux impliqués dans l'explication des maladies. (5) Enfin, ajoutons que cette stratégie épistémologique a également des conséquences éthiques discutables. Maintenir une distinction artificielle entre des maladies génétiques et des maladies non génétiques revient à accréditer la thèse de l'exceptionnalisme génétique qui considère que les maladies génétiques posent des problèmes éthiques spécifiques. Or, avec le développement des techniques de séquençage de génome entier à un prix de plus en plus modique et à une vitesse toujours plus rapide, c'est l'ensemble des maladies et non pas un petit groupe de maladies génétiques, qui vont être concernées par le développement des tests de diagnostic et de dépistage génétiques et par les potentiels problèmes éthiques que ces développements technologiques peuvent poser.

1.2. Une multiplicité de théories génétiques possibles

Pour résoudre le paradoxe de la génétique médicale contemporaine, nous avons proposé la stratégie inverse : il s'agit d'abandonner le concept de maladie génétique et de prendre au sérieux la généticisation des maladies en faisant l'hypothèse qu'elle révèle l'élaboration d'une explication du rôle commun des gènes dans l'apparition et le développement des maladies, ce que nous avons appelé « une théorie génétique de la maladie ».

Pour éviter qu'une théorie génétique soit une explication triviale ou une explication génocentriste des maladies, nous avons établi au chapitre 3 la condition de possibilité d'une théorie génétique : il ne peut y avoir de théorie génétique en dehors d'une explication interactionniste co-constructionniste des maladies, qui ne donne pas de prédominance causale aux gènes et qui permette d'identifier la matrice causale complexe de l'explication des maladies.

Au chapitre 4, nous avons établi les critères d'identification d'une théorie génétique : une théorie génétique *a minima*, est une théorie médicale partielle qui est centrée sur le rôle commun des gènes et qui déploie au moins un type d'explication

(physiopathologique, épidémiologique ou évolutionnaire) d'au moins une cible explicative (causes, évolution, symptômes, traitement) d'une maladie, d'une classe de maladies ou de toutes les maladies. Nous avons également identifié trois critères d'évaluation d'une théorie génétique : l'explication de la variabilité interindividuelle, de la variabilité intra-individuelle ou de la comorbidité sont trois cibles explicatives qu'une théorie génétique ne doit pas nécessairement expliquer mais qui augmentent la précision de la théorie et son pouvoir explicatif. Nous avons ainsi établi qu'il pouvait exister en principe une multiplicité de théories génétiques potentielles à partir de la combinaison de ces différents critères, selon le niveau d'explication, le type de cibles explicatives, le type d'explication mobilisé et le type de critères d'évaluation mis en œuvre (voir Figure 65) .

Cependant, étant donné que notre objectif était de comprendre la généticisation de toutes les maladies et ses conséquences sur notre conception de la maladie, nous avons proposé de nous concentrer sur deux formes particulières de théories génétiques générales, c'est-à-dire des théories qui visent à expliquer le rôle commun des gènes dans toutes les maladies, et pas seulement dans une classe de maladie ou dans une maladie-type : la *théorie génétique des maladies* et la *théorie génétique de la maladie*.

Nous avons ainsi transformé notre hypothèse initiale en une alternative : ou bien la généticisation révèle *une théorie génétique des maladies*, c'est-à-dire un ensemble hétérogène de théories génétiques régionales, chacune étant spécifique d'une classe de maladies, ou bien la généticisation révèle *une théorie génétique de la maladie*, c'est-à-dire une théorie qui repose sur une définition de la maladie qui unifie le rôle commun des gènes, à partir de laquelle toute maladie pourrait être instanciée et qui serait susceptible de renouveler en profondeur la façon dont nous classons les maladies.

Condition de possibilité		Théorie partielle centrée sur le rôle des gènes au sein d'une explication interactionniste co-constructionniste des maladies
Critères d'identification	**Objet**	La maladie
		Une classe de maladie
		Une maladie-type
	Niveau	Théorie génétique générale : *théorie génétique de la maladie* ou *théorie génétique des maladies*
		Théorie génétique régionale
		Théorie génétique d'une maladie
	Cibles explicatives	Causes
		Symptômes
		Évolution
		Traitement
	Type d'explication	Explication épidémiologique
		Explication évolutionnaire
		Explication physiopathologique
Critères d'évaluation		Variabilité interindividuelle
		Variabilité intra-individuelle
		Comorbidité

Figure 65 : Condition de possibilité, critères d'identification et critères d'évaluation d'une théorie génétique. On peut parler d'une théorie génétique *a minima* à partir du moment où une théorie centrée sur le rôle des gènes dans une explication interactionniste co-constructionniste des maladies. Une théorie génétique mobilise au moins un type d'explication (épidémiologique, évolutionnaire, physiopathologique) pour rendre compte au moins d'une classe explicative (causes, symptômes, évolution, traitement) d'un objet-maladie (la maladie, l'ensemble des maladies, une classe de maladie, une maladie-type). Une théorie qui rend compte de l'explication de toutes les maladies est une théorie génétique générale et peut être soit une *théorie génétique des maladies*, un ensemble hétérogène de théories génétiques régionales, chaque théorie génétique étant spécifique d'une classe de maladie, soit une *théorie génétique de la maladie*, c'est-à-dire une théorie générale qui repose sur une définition générale de la maladie qui unifie le rôle causal commun des gènes. Trois critères (explication de la variabilité interindividuelle, de la variabilité intra-individuelle, de la comorbidité) permettent d'évaluer le degré de précision d'une théorie génétique donnée à un niveau donné.

1.3. Une multiplicité de théories génétiques de fait

Pour tester cette alternative, nous avons identifié dans la littérature biomédicale contemporaine deux exemples de théories génétiques : d'une part la théorie génétique des maladies infectieuses, d'autre part la théorie génétique au sein de la médecine des réseaux.

La théorie génétique des maladies infectieuses a été développée par l'équipe de Casanova et Abel pour expliquer la variabilité interindividuelle aux infections, qui était jusqu'ici non expliquée ou imparfaitement expliquée par les postulats de Koch et par ce que Casanova et Abel appellent la « théorie microbienne », la « théorie immunologique » et la « théorie écologique » des maladies infectieuses.

La théorie génétique des maladies infectieuses est une théorie régionale, spécifique à la classe des maladies infectieuses. Elle contribue à l'explication des quatre cibles explicatives (causes, symptômes, évolution, traitement) des maladies infectieuses en déployant une explication physiopathologique qui identifie quatre mécanismes génétiques spécifiques. Elle permet d'expliquer la variabilité interindividuelle clinique aux maladies infectieuses, en particulier le taux élevé de mortalité infantile dans les infections primaires, de compléter l'explication de la variabilité interindividuelle des adultes aux infections secondaires et aux réactivations, et enfin d'expliquer le problème des infections inapparentes.

Si la théorie génétique des maladies infectieuses constitue une explication co-constructionniste passionnante des maladies infectieuses, elle présente cependant quelques limites. Elle repose en effet sur l'identification de mécanismes hérités de la génétique classique (prédisposition mendélienne monogénique, prédisposition polygénique), dont nous avons montré qu'ils étaient en partie remis en question aujourd'hui. Elle intègre imparfaitement le rôle de certains facteurs causaux dans l'explication des maladies infectieuses, comme le rôle de la génétique du parasite et le rôle du microbiome. Enfin, elle ne nous dit rien de ce que nous avons appelé les « cas-limites » de maladies infectieuses et qui recouvrent les cancers viro-induits, ainsi que les maladies auto-immunes et les maladies chroniques liées à des infections.

La médecine des réseaux est née de la conjonction d'une nouvelle méthode de recherche (la biologie et la médecine des systèmes), d'un nouveau concept (le concept

de gènes humains de la maladie) et d'une nouvelle boîte à outils (la théorie des réseaux). C'est une théorie générale de la maladie, qui définit la maladie comme la conséquence de perturbations d'un module fonctionnel et au sein de laquelle nous avons isolé une théorie génétique de la maladie.

Cette théorie génétique de la maladie permet à travers une explication topologique d'expliquer l'origine génétique commune des maladies et le phénomène de comorbidité, de redéfinir la distinction entre maladies monogéniques et maladies polygéniques comme une distinction entre des maladies dont les modules impliquent une redondance fonctionnelle plus ou moins élevée et qui sont par conséquent plus ou moins robustes aux différentes perturbations (stochastiques, épigénétiques, génétiques, environnementales).

Cette théorie génétique de la maladie est enfin remarquable par sa dimension heuristique : elle permet de découvrir de nouveaux gènes de la maladie, de mettre en place de nouvelles cibles thérapeutiques et elle incite à proposer des explications mécanistes pour comprendre le fonctionnement des modules de la maladie.

1.4. La nécessaire coexistence de théories génétiques variées

Ce travail aboutit donc à deux résultats. D'une part, nous avons démontré qu'en principe, il existe potentiellement une infinité de théories génétiques possibles selon la façon dont on combine le niveau, le type d'explication, le type de cible explicative, et les critères d'évaluation mis en œuvre. D'autre part, nous avons montré qu'il existe de fait au moins deux exemples de théories génétiques : une théorie génétique régionale qui peut servir de modèle à une théorie génétique générale des maladies et une théorie génétique au sein de la médecine des réseaux qui peut servir de modèle à une théorie génétique générale de la maladie.

Ces deux résultats ne sont pas contradictoires, mais ils semblent empêcher la résolution de notre alternative : la généticisation relève-t-elle d'une théorie génétique des maladies ou d'une théorie génétique de la maladie ? Pour le dire différemment, la coexistence de principe et de fait de différents types de théories génétiques pose un problème d'unité théorique : les différents types de théories génétiques sont-ils dans un rapport de concurrence ou de complémentarité ?

Loin d'être problématique, la coexistence de théories génétiques variées nous apparaît nécessaire pour trois raisons.

D'une part, comme nous l'avons montré avec l'exemple des maladies infectieuses, ces deux types de théories répondent à des questions différentes, avec des formes d'explications (topologique, mécaniste) qui permettent d'éclairer différents aspects d'une même maladie ou d'une même classe de maladie (comorbidité, variabilité interindividuelle). C'est donc d'abord pour des raisons pragmatiques que la coexistence de différents types de théories génétiques nous paraît nécessaire.

D'autre part, la coexistence de théories génétiques nous paraît souhaitable d'un point de vue institutionnel, du point de vue de la recherche comme du point de vue de la clinique. Du point de vue de l'organisation scientifique de la recherche biomédicale, la nécessité d'approches complémentaires est soulignée aussi bien par les partisans des approches systémiques que par les partisans d'approches plus traditionnelles de biologie moléculaire. Du point de vue de l'organisation clinique, il est évident que l'organisation du système de santé actuel dépend en partie de notre classification des maladies avec des spécialités médicales qui correspondent à des classes de maladie. Tant que la théorie génétique de la médecine des réseaux n'a pas évolué au point de proposer une nouvelle classification générale des maladies, il nous semble que la coexistence de théories génétiques différentes de la maladie permet à la fois d'incorporer la génétique médicale dans chaque spécialité et d'identifier des caractéristiques de la génétique médicale des maladies infectieuses par exemple, tout en permettant d'élaborer des approches plus décloisonnées de certaines maladies (cancers et maladies infectieuses par exemple).

Enfin, une dernière raison nous paraît justifier la nécessité de faire appel à des théories génétiques variées pour résoudre le paradoxe de la génétique contemporaine : c'est le développement récent de trois programmes de recherche qui vont probablement complexifier encore notre compréhension de la génétique des maladies.

2. Trois autres sources de complexité pour la génétique médicale

Lorsque nous avons exposé les difficultés scientifiques posées par le concept de maladie génétique, nous nous sommes concentrée sur la description de phénomènes biologiques et de travaux de génétique qui concernent principalement l'ADN nucléaire ou mitochondrial codant humain. Cependant, trois autres domaines de recherche, qui ont pris une ampleur considérable ces dix dernières années, sont susceptibles de bouleverser encore davantage les contours de la génétique médicale. Il s'agit (1) des fonctions du génome non codant, (2) du rôle du microbiome et (3) de la place de l'épigénétique dans l'explication des maladies humaines.

2.1. Les fonctions du génome non codant

Nous avons déjà mentionné que seule une très petite partie du génome était codante, c'est-à-dire que seul 1,5% du génome produit des protéines. L'ADN non-codant a ainsi été longtemps considéré comme de « l'ADN poubelle » (*junk DNA*), qui serait un reliquat de l'évolution et entre autres du processus de duplication génétique (Ohno, 1970). Mais le projet ENCODE (*Encyclopedia of DNA elements*), lancé en 2003 et dont l'objectif était de cartographier tous les éléments fonctionnels du génome, a révélé, dans une analyse détaillée de 2012, que 80% de l'ADN dit non-codant serait tout de même fonctionnel, au sens où 80% de l'ADN non codant serait impliqué dans une activité biochimique cellulaire[62] et jouerait un rôle décisif en termes de régulation de l'expression de l'ADN codant (ENCODE Project Consortium *et al.*, 2012).

Le génome non codant comporte ainsi des séquences d'ADN non codantes mais régulatrices (les promoteurs, les *enhancers*, les *silencers*, etc.), des éléments mobiles (comme les transposons), des éléments viraux comme des rétrovirus endogènes, des séquences qui ont perdu leur capacité à coder des protéines comme les pseudogènes, mais aussi des ARN non codants qui sont des molécules ARN fonctionnelles qui ne codent pas des protéines. Il existe toutes sortes d'ARN non codants (Théry, 2013) dont le rôle pathogénique a été mis en évidence dans de nombreuses maladies (Esteller,

[62] Pour une analyse de la controverse qui a accompagné cette annonce et qui concerne la définition du concept de fonction, on se reportera à (Germain *et al.,* 2014)

2011 ; Shi *et al.,* 2013 ; Ward et Kellis, 2012) et dont la Figure 66 donne quelques exemples aussi différents que la transition épithélio-mésenchymateuse de tous les cancers, la maladie d'Alzheimer et le syndrome de Prader-Willis. De façon plus générale, l'ensemble des éléments fonctionnels du génome non codant est susceptible d'avoir des implications considérables sur notre compréhension de la génétique des maladies.

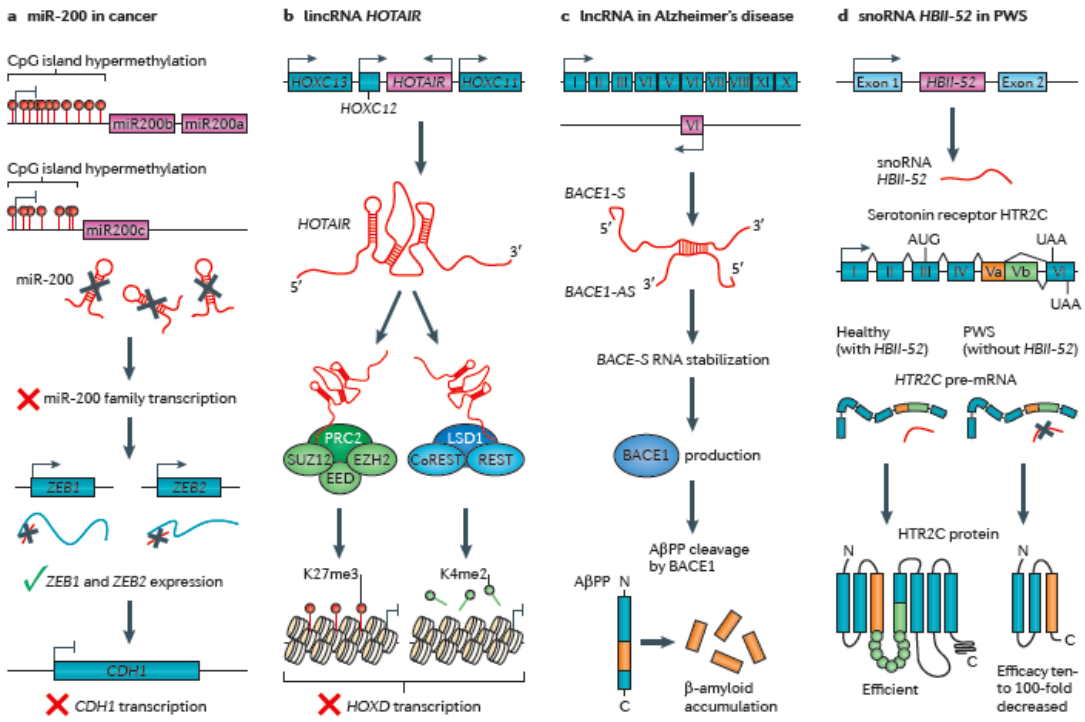


Figure 66 : Exemples du rôle des ARN non codants dans la pathogénèse. (Esteller, 2011, p. 865) Figure a : miR-200 est un exemple de microARN dont le rôle dans la formation des tumeurs a été bien caractérisé. Des altérations dans la régulation épigénétique (hyperméthylation des îlots CpG) de la famille miR-200 sont impliquées dans la transition épithélio-mésenchymateuse (TEM) qui donne des propriétés invasives aux cellules cancéreuses et est impliquée dans le développement de métastases. Figure b : L'ARN non codant intergénique long (lincRNA) HOTAIR est surexprimé dans les cellules cancéreuses, entrainant un remodelage de la chromatine qui augmente les capacités invasives des cellules cancéreuses. Figure c : La maladie d'Alzheimer est caractérisée par une accumulation de plaques de béta-amyloïde dans le cerveau. Le long ARN non codant antisens BACE1-AS permettrait de stabiliser l'ARN messager de BACE1 qui est impliqué dans le clivage anormal de la protéine APP, qui cause l'accumulation de la béta-amyloïde caractéristique de la maladie d'Alzheimer. Figure d : la perte du petit ARN nucléolaire snoRNA aurait des conséquences importantes dans le syndrome de Prader-Willi (PWS).

2.2. Le rôle du microbiome

Par ailleurs, avec l'avènement de techniques de séquençage toujours plus rapides, il est maintenant possible d'explorer le microbiome humain, c'est-à-dire l'ensemble des génomes des microorganismes commensaux, symbiotiques ou pathogènes qui habitent le corps humain. Le microbiome humain varie grandement d'un individu à un autre et au fil de la vie d'un même individu. De nombreux facteurs sont impliqués dans sa composition (rôle de l'environnement maternel, du mode de naissance par voie basse ou par césarienne, de l'exposition aux pathogènes pendant les premières années de vie, de la nutrition, etc.). Le Projet Microbiome Humain, lancé par le NIH en 2008 (Badger *et al.,* 2011), a permis d'évaluer qu'un être humain cohabite avec environ 1 000 000 milliards de micro-organismes, (soit dix fois plus de cellules microbiennes que de cellules humaines), qui représentent environ 10 000 espèces microbiennes différentes et contiennent 8 millions de gènes uniques codant des protéines (contre environ 20 000 gènes humains).

Cette exploration nous amène, pour reprendre les termes de John Dupré, à dépasser l'idée communément acceptée selon laquelle les êtres humains sont des organismes composés de lignées cellulaires qui se sont différenciées à partir d'une seule cellule au génome unique (Dupré, 2011). Les génomes d'environ 10^{14} micro-organismes cohabitent dans un même organisme humain et ont une influence considérable sur l'apparition ou la prévention des maladies, que ce soit par les fonctions métaboliques dans lesquelles ils sont impliqués (Nicholson *et al.,* 2012) ou par leur modulation de notre système immunitaire (Hooper *et al.,* 2012). On parle ainsi de métagénome[63] pour désigner l'ensemble du génome d'un hôte et des génomes des micro-organismes qui vivent sur ou dans un hôte à un moment donné. Le métagénome est ainsi composé d'éléments transitoires (par exemple, lors de l'infection par un pathogène) et d'éléments plus persistants (par exemple, l'infection par des virus eucaryotes latents ou la présence de bactéries commensales). La métagénomique désigne ainsi une approche qui ne considère pas seulement les gènes isolés d'un hôte

[63] Le terme de métagénome peut être parfois utilisé un peu différemment. Pour une mise en perspective historique et une analyse philosophique des implications de ce concept, on peut se reporter à (Dupré et O'Malley, 2007).

ou les gènes isolés d'un micro-organisme mais qui prend en compte l'ensemble du métagénome et cherche à comprendre la façon dont le génome de l'hôte et le génome du microbiome interagissent pour moduler l'apparition des maladies.

L'implication du microbiome à de multiples étapes de l'apparition de certaines maladies, comme le diabète de type 1, la maladie de Crohn et la rectocolite-hémorragique est telle (Figure 67) que certains auteurs ont proposé de définir ces maladies comme des maladies « métagénétiques » plutôt que comme des maladies génétiques (Virgin et Todd, 2011).

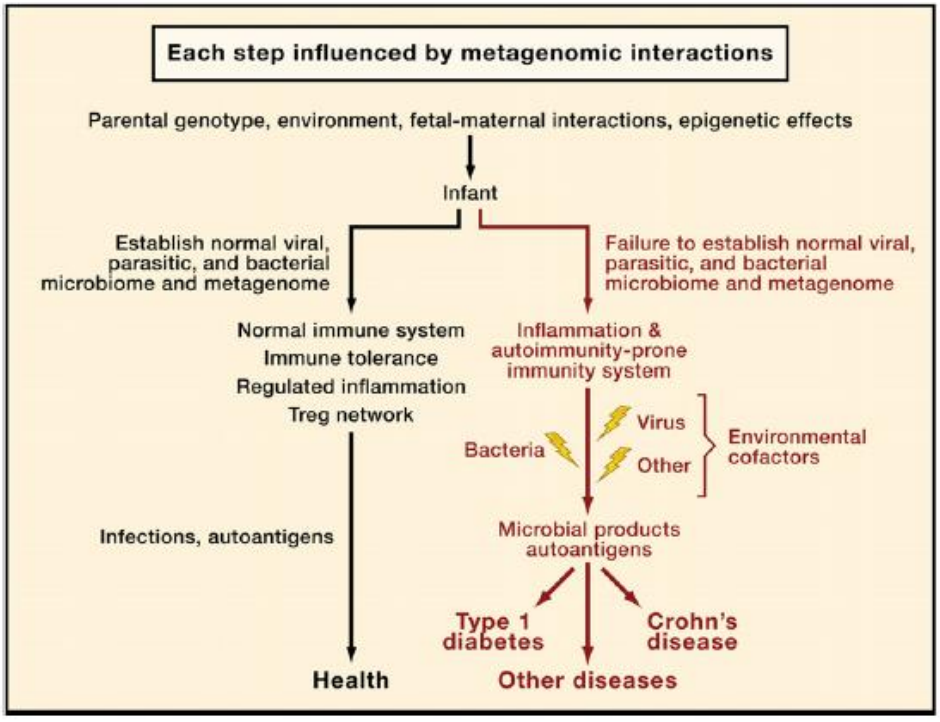


Figure 67 : Une vision métagénomique du développement de l'immunité normale et des maladies auto-immunes (Virgin et Todd, 2011, p. 50). Le microbiome s'acquiert dans les premières années de vie. A gauche, l'acquisition d'un microbiome et d'un métagénome normal permet d'établir un système immunitaire efficient et mène à la santé. A droite, l'acquisition d'un microbiome et d'un métagénome perturbés entrainent un système immunitaire qui a une tendance à l'inflammation et à l'auto-immunité. L'exposition à certains cofacteurs environnementaux (infections virales ou bactériennes, autres facteurs environnements) entraine la production d'autoantigènes microbiens qui peuvent être impliqués dans certaines maladies comme le diabète de type 1 ou la maladie de Crohn, considérées comme des maladies métagénétiques.

2.3. La place de l'épigénétique

L'épigénétique est un autre domaine de recherche extrêmement dynamique dont les implications en termes de génétique médicale se sont multipliées, au point

que certains parlent d'un nouveau « principe unifiant » (Petronis, 2010) ou d'un « nouveau paradigme » (Ferguson-Smith, 2011), et qualifient certaines maladies qui impliquent des phénomènes épigénétiques d'empreinte parentale par exemple de « maladies épigénétiques ».

Comme nous l'avons mentionné ailleurs, l'épigénétique couvre un ensemble très large de modifications de l'expression du génome et de sa régulation, qui ont la particularité de ne pas modifier la séquence nucléotidique mais qui peuvent être transmissibles d'une génération à une autre, voire sur plusieurs générations et dont certaines sont réversibles. Les trois familles principales de modifications épigénétiques sont la méthylation de l'ADN, les modifications des histones et le remodelage de la chromatine. Ces trois familles de modifications épigénétiques sont susceptibles d'être impliquées dans l'apparition des maladies (Feinberg, 2007 ; Jiang *et al.*, 2004 ; Portela et Esteller, 2010). L'épigénome désigne l'ensemble des modifications épigénétiques du génome d'un organisme et peut être modifié par de nombreux facteurs environnementaux : c'est donc un moteur fondamental de la plasticité phénotypique qui permet aux cellules de modifier leur comportement en fonction des signaux internes et externes qu'elles reçoivent. C'est la dérégulation de cette capacité d'adaptation de l'expression génomique qui serait impliquée dans de nombreuses maladies (Feinberg, 2007).

Le cancer est une des classes de maladie dont les modifications épigénétiques ont été les plus étudiées. Tous les cancers sont en effet caractérisés par un ensemble de modifications épigénétiques communes impliquant, entre autres, une hypométhylation globale du génome, une hyperméthylation des îlots CpG dans les régions promotrices des gènes suppresseurs de tumeurs, des modifications de la structure des histones et en particulier une perte globale de l'histone H4 mono-acétylé et tri-acétylé. Certains chercheurs ont ainsi proposé de redéfinir le cancer comme un *« épigénome dérégulé qui permet la croissance tumorale aux dépens de l'hôte »* (Timp et Feinberg, 2013) : un ensemble de mutations génétiques et épigénétiques modifieraient l'épigénome, entrainant une incapacité de la cellule à maintenir son intégrité épigénétique. Cette incapacité à maintenir l'intégrité de l'épigénome entraine une variabilité épigénétique accrue, caractéristique de tous les cancers, qui conduit à une dérive phénotypique des cellules cancéreuses et à l'apparition des propriétés qui

sont souvent considérées comme étant essentielles au développement des cancers (Hanahan et Weinberg, 2000, 2011) : le maintien de signaux de croissance, l'insensibilité aux signaux suppresseurs de croissances, la résistance à l'apoptose, la capacité de se répliquer indéfiniment, l'induction de l'angiogenèse, la capacité à envahir des tissus et à former des métastases, la reprogrammation du métabolisme énergétique et la capacité à éviter la destruction par le système immunitaire (voir Figure 68).

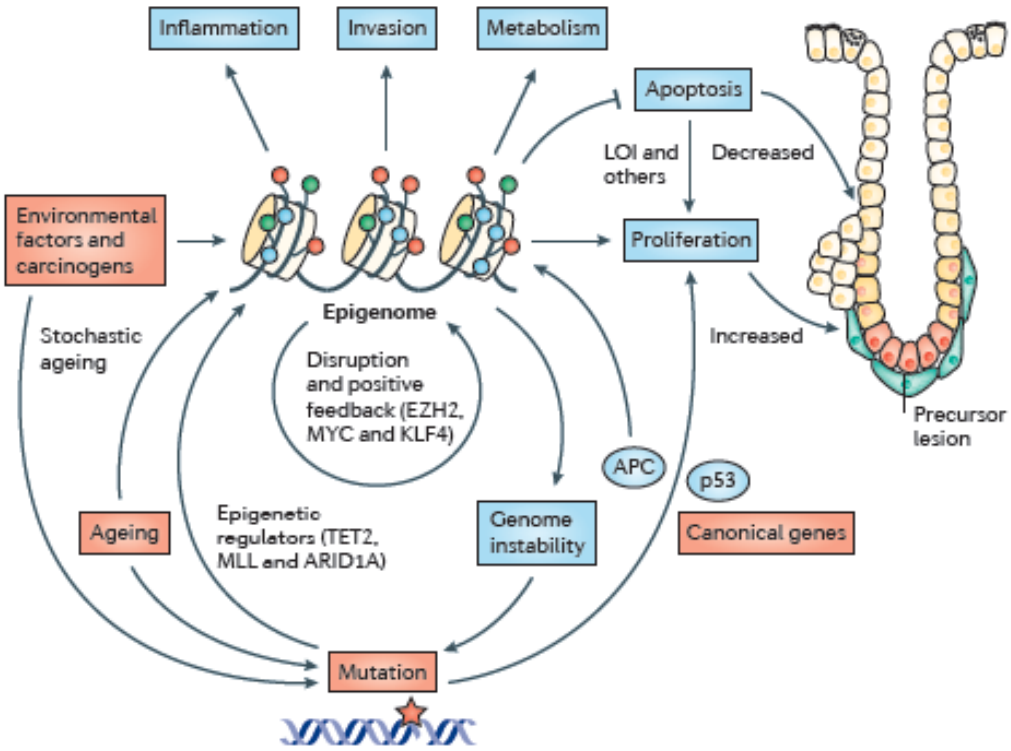


Figure 68 : Collaboration des modifications épigénétiques et des mutations dans le développement des propriétés des cellules cancéreuses (Timp et Feinberg, 2013, p. 507). Les facteurs environnementaux (carcinogènes, régime alimentaire) entrainent une reprogrammation épigénétique. L'épigénome accumule également des dommages de façon stochastique et au cours du vieillissement. La machinerie qui maintient l'intégrité génétique peut être perturbée soit par des mutations génétiques, soit par des modifications épigénétiques qui exercent un rétrocontrôle positif. La perturbation du maintien de l'intégrité épigénétique conduit à une perte de la régulation épigénétique et à une dérive phénotypique stochastique, suivie d'une sélection des cellules avec une forte croissance aux dépens des autres cellules. Certaines modifications épigénétiques, comme le décalage des zones de méthylations sur les îlots CpG entrainent des changements métaboliques et augmentent les capacités de prolifération. D'autres, comme les blocs hypométhylés, augmentent le potentiel invasif. D'autres encore modifient directement l'équilibre entre l'apoptose et la prolifération. Des mutations génétiques canoniques, comme celle de TP53 (qui encode le gène p53, muté dans 2/3 des cancers) peuvent directement affecter les différentes propriétés des cellules cancéreuses mais peuvent aussi mener à la dérégulation de l'épigénome. De la même manière, des perturbations épigénétiques, comme l'hypométhylation globale ou le l'hperméthylation des îlots CpG peuvent entrainer à leur tour des réarrangements chromosomiques et des mutations génétiques.

Pour résumer, les problèmes scientifiques posés par le concept de maladie génétique que nous avons principalement étudiés dans ce travail concernent avant tout les 1,5% de notre génome codant. Mais la génétique médicale devra également intégrer dans les prochaines années l'étude des 98,5% de l'ADN non codant de notre génome, les 10^{14} génomes du microbiome humain et l'ensemble des modifications épigénétiques qui sont susceptibles de modifier l'expression du génome codant, du génome non codant et du microbiome. Cette constatation nous paraît justifier d'autant plus la nécessité d'élaborer des théories génétiques qui permettent d'unifier le rôle des gènes dans l'explication des maladies. Mais elle justifie aussi l'existence d'une multiplicité de théories génétiques : il paraîtrait naïf d'imaginer qu'une même théorie aurait la capacité à la fois d'expliquer l'ensemble de ces phénomènes pour toutes les maladies et de répondre à tous les contextes (recherche biomédicale, clinique, thérapeutique, etc.) où l'explication du rôle des gènes dans une maladie est nécessaire.

3. Le concept de théorie médicale et la philosophie de la médecine

Nous voudrions terminer par quelques remarques plus générales sur la portée épistémologique notre travail.

L'épistémologie de la médecine a été longtemps marquée par le débat entre naturalistes et normativistes sur la définition des concepts de santé et de maladie. Si ces débats sont essentiels, ils ne sont pas les seuls à devoir être abordés dans le cadre de l'épistémologie de la médecine. Ces dernières années, d'autres questions ont émergé, qui concernent le statut de la médecine des preuves dans l'élaboration et la structuration du savoir biomédical. En revanche, à l'exception de quelques rares auteurs, il semble que les théories médicales suscitent un scepticisme persistant, tant du point de vue de leur contenu que de celui de leur utilité. Ce scepticisme (voir Annexe 2) est souvent lié à l'idée que la médecine est avant tout un savoir pratique, dont la constitution ne peut être guidée que par des objectifs thérapeutiques et a souvent pour conséquence de séparer d'un côté les sciences biologiques et de l'autre, la pratique médicale.

Ce travail est une réponse à ce scepticisme et à cette partition. Pour résoudre le paradoxe de la génétique médicale, nous avons élaboré un concept de théorie génétique et de façon plus générale un concept de théorie médicale. Ce concept de théorie médicale a l'avantage de pouvoir s'adapter à la diversité des théories qui existent en médecine, des « petits bouts d'explication causale » – qu'un grand sceptique vis-à-vis des théories médicales comme Harold Kincaid reconnaît exister dans l'explication des maladies-type (Kincaid, 2008) – aux théories générales de la maladie, comme celle de la médecine des réseaux. Surtout, ce concept de théorie médicale permet au clinicien de passer d'une explication générale de la maladie, d'une classe de maladie ou d'une maladie-type à l'explication clinique de la maladie incarnée d'un patient particulier. Accepter que la philosophie de la médecine doive se préoccuper de l'existence des théories médicales, les mettre à jour et évaluer leur pouvoir explicatif ne nous paraît donc en aucun cas contradictoire avec le postulat que l'objectif du médecin est avant tout de guérir ses patients. Mais il nous semble, comme le fait remarquer Norbert Paul (Paul, 1998), qu'un « *hiatus theoreticus* » ne cesse de grandir entre la recherche biomédicale d'une part et la pratique clinique d'autre part, entre l'accumulation de connaissances théoriques en médecine et leur mise en pratique dans le cadre de la clinique. En proposant un concept de théorie médicale *a minima*, nous espérons proposer un outil conceptuel utile à la résolution de cet *hiatus theoreticus*.

Annexes

Annexe 1 : (Chapitre 3) – Méthodes d'identification des variants alléliques – des maladies monogéniques au cauchemar des maladies complexes

Les années 1980 et 1990 ont été une période faste d'élucidation des composants génétiques des maladies mendéliennes, principalement grâce au clonage positionnel aussi appelée « génétique inverse » (Jordan, 1988, 2006). En effet, la génétique dite « classique » s'appuie sur une approche aussi appelée « physiologique » : partant de ce que l'on connaît de la physiologie de la maladie (symptômes, fonction des protéines impliquées, etc.), on cherche à retrouver le gène ou les gènes porteur(s) des mutations impliquées dans la maladie.

Au contraire, la génétique inverse ou clonage positionnel cherche à identifier le gène en identifiant sa position dans le génome. Cette méthode s'appuie sur l'analyse de liaison, c'est-à-dire sur le fait que des allèles dont les locus sont proches (par exemple, des allèles qui sont sur le même chromosome) ont tendance à coségréger, c'est-à-dire à être transmis ensemble à la génération suivante lors de la reproduction. Pour identifier les variants alléliques responsables d'une maladie donnée grâce au clonage positionnel, il faut donc plusieurs conditions : (1) il est nécessaire de connaître le mode de transmission de la maladie (mendélien autosomique dominant, mendélien autosomique récessif, etc.) – ce qui s'établit généralement par une analyse de ségrégation (on explore l'arbre génétique de familles atteintes par la maladie pour déduire son mode de transmission) (2) il est nécessaire de disposer de marqueurs qui jalonnent l'ensemble de la cartographie du génome : cette condition est maintenant remplie puisqu'on dispose de différents types de marqueurs tels que les polymorphismes de restrictions de fragments polymorphes (ou RFPL), les locus de caractère quantitatif (ou QTL) et plus récemment les polymorphismes nucléotidiques simples (SNP ou snip), (3) il est préférable d'avoir un phénotype malade clairement identifiable (plus le diagnostic est sans ambiguïté, et plus on évite à la fois les erreurs sur les hypothèses du mode de transmission et les problèmes d'échantillonnage).

Le clonage positionnel se fait alors en quatre étapes : (1) d'abord on réalise sur de grandes familles de patients atteints une analyse de liaison génétique, c'est-à-dire qu'on teste pour plusieurs marqueurs la probabilité que ces marqueurs coségrégent avec l'allèle morbide selon la méthode des *lods-scores*. Une fois qu'on a repéré ce type de marqueur (dont la localisation sur le génome est connu), on en déduit que l'allèle morbide doit se trouver dans le voisinage de ce marqueur, (2) il faut alors réaliser une carte génétique de la sous-région chromosomique identifiée en utilisant de nouveaux marqueurs (3) puis identifier tous les gènes de cette région (4) et enfin rechercher des mutations dans ces gènes et identifier le gène responsable. Cette méthode est donc relativement longue mais a permis par exemple la découverte de l'implication du gène CFTR dans la mucoviscidose en 1987 ou la découverte du gène HD dans la maladie de Huntington.

Par ailleurs, des approches mixtes entre clonage positionnel et clonage fonctionnel peuvent être utilisées telles que l'approche dite « gène-candidat » ou l'approche dite « candidat positionnel ». L'approche dite « gène-candidat » est utilisée lorsqu'un gène a été identifié chez une autre espèce comme responsable d'une affection similaire à la maladie d'intérêt. On peut alors rechercher dans le génome humain une séquence similaire et identifier les mutations d'intérêt. L'approche dite « candidat positionnel » intervient lorsque l'on peut combiner une analyse de liaison génétique à des « gènes candidats » (gènes dont le positionnement et ou la fonction sont connus et qui sont dans la région d'intérêt). La Figure 69 ci-dessous récapitule les différentes techniques d'identification des gènes évoquées jusqu'ici.

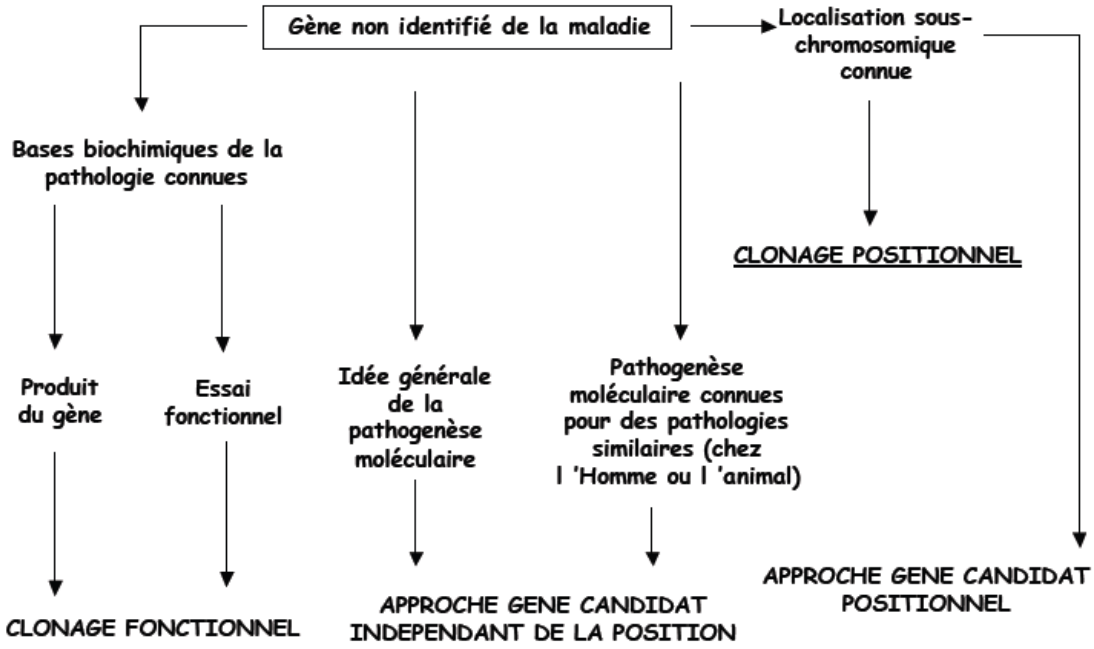

Gène non identifié de la maladie
Localisation sous-chromosomique connue
Bases biochimiques de la pathologie connues
CLONAGE POSITIONNEL
Produit du gène
Essai fonctionnel
Idée générale de la pathogenèse moléculaire
Pathogenèse moléculaire connues pour des pathologies similaires (chez l 'Homme ou l 'animal)
APPROCHE GENE CANDIDAT POSITIONNEL
CLONAGE FONCTIONNEL
APPROCHE GENE CANDIDAT INDEPENDANT DE LA POSITION

Figure 69 : Les différentes méthodes de "chasse aux gènes" dans les maladies mendéliennes

Qu'en est-il des maladies polygéniques ou complexes ? Peut-on utiliser les mêmes techniques et en particulier, celles du clonage positionnel qui ont connu un succès colossal pour les maladies mendéliennes ? L'utilisation des techniques du clonage positionnel est certes possible, mais beaucoup plus incertaine et inefficace dans le cas des maladies polygéniques (Botstein et Risch, 2003) – au point que certains auteurs parlent de l'élucidation de la génétique des maladies complexes comme un « cauchemar du généticien » (Feingold, 2005). Plusieurs raisons expliquent cette difficulté : la première est que l'analyse de liaison génétique, telle qu'elle est pratiquée dans le cas des maladies mendéliennes est dite « paramétrique » : cela signifie qu'on connaît un certain nombre de paramètres comme le mode de transmission de la maladie, qui sont nécessaires au calcul de la vraisemblance des liaisons génétiques entre un marqueur donnée et la maladie d'intérêt dans l'analyse des grandes familles. Ces données ne sont pas disponibles pour les maladies polygéniques, qui, comme nous l'avons rappelé dans le chapitre 1, mettent souvent en jeu soit un mode de transmission dit polygénique classique (un très grand ensemble d'allèles à effet mineur sont impliqués dans la maladie), soit un mode de transmission plus complexe (maladies polygéniques à gène majeur, maladies oligogéniques), etc. Par ailleurs, il est plus difficile d'obtenir des grandes familles pour de telles analyses de liaison : les diagnostics sont parfois ambigus, il n'est pas toujours facile de distinguer au sein de la

famille les cas qui relèvent d'une transmission héréditaire de ceux qui relèvent de mutations sporadiques (c'est le cas par exemple du cancer du sein). Le fait que les allèles impliqués dans la plupart des maladies communes aient une pénétrance incomplète risque également de fausser ces études, puisqu'on exclut de l'étude de ces grandes familles les patients sains qui sont peut-être pourtant porteurs d'un allèle de susceptibilité à la maladie qui ne s'exprimerait pas. Enfin, des phénomènes comme l'hétérogénéité génétique (des gènes différents peuvent donner la même maladie) ou l'existence de gènes modificateurs viennent compliquer l'analyse, plus souvent que dans le cadre des maladies monogéniques. Pour pallier ces inconvénients, une variante de l'analyse de liaison paramétrique utilisée pour les maladies mendéliennes a été développée pour les maladies polygéniques : les analyses de liaison non paramétriques, telles que la méthode des paires de germains atteints. Ces méthodes ont cependant leurs limites (Boehnke, 1994). Outre qu'elles reposent sur une estimation simplifiée des paramètres impliqués dans les études de liaison, elles nécessitent également des échantillons très larges, particulièrement lorsqu'il s'agit de rechercher des variants alléliques communs ayant un effet mineur :

> « En effet, il y a maintenant plusieurs exemples de variants communs qui contribuent aux maladies communes, la plupart d'entre eux multiplient le risque de la maladie par deux (parfois moins) lorsqu'ils sont examinés dans de larges populations. Il est clair que de tels allèles de susceptibilité à la maladie qui sont bien établis n'auraient pas pu être détectés par l'analyse de liaison. Par exemple, le variant Pro12A1 dans le gène du récepteur gamma activé par les proliférateurs de peroxysome (gène PPARG), qui affecte le risque de développer le diabète de type 2, aurait nécessité pour être détecté par des analyses de liaison d'utiliser un échantillon de plus d'un million de paires de germains atteints. » (Hirschhorn et Daly, 2005, 96)

Pour étudier la génétique des maladies complexes, des alternatives à l'étude de liaison génétique ont donc été proposées. La recherche des gènes candidats que nous avons déjà évoquée pour les maladies mendéliennes, a bien sûr été utilisée. On utilise

alors des informations sur la physiopathologie de la maladie ou sur des gènes similaires identifiés dans des pathologies comparables chez d'autres espèces pour comparer des séquences entières d'une même région ou d'un même gène entre des individus atteints par la maladie et des individus sains. Cependant, même avec la diminution des coûts et l'augmentation de la rapidité des techniques de séquençage, ce type d'étude reste coûteux – ce qui explique que les chercheurs impliqués se limitent généralement à séquencer un ou deux gènes d'intérêt, au risque de ne pas parvenir à tirer des conclusions significatives. Une autre méthode pour adapter les études de gènes candidats aux maladies complexes et qui est moins coûteuse que le reséquençage de gènes candidats est l'étude d'association de gènes candidats. Les études d'association en général utilisent le principe du déséquilibre de liaison. Qu'est-ce que le déséquilibre de liaison ? Le déséquilibre de liaison mesure la fréquence avec laquelle les allèles de deux polymorphismes se retrouvent sur le même chromosome. Il est affecté par la recombinaison méiotique et diminue proportionnellement à la distance séparant les polymorphismes. Quand il existe un déséquilibre de liaison entre deux allèles, cela signifie qu'il faudra encore un grand nombre de recombinaisons méiotiques (c'est à dire un grand nombre de générations) pour que les deux polymorphismes en question soient en équilibre de liaison, c'est-à-dire pour qu'il y ait la même probabilité pour que les deux allèles coségrégent que pour qu'ils ségrégent séparément durant la méiose. Les études d'association différent significativement des études de liaison génétique : alors que les études de liaison portent sur l'étude de liaisons génétiques au sein de grandes familles de patients atteints par une maladie donnée, les études d'association génétique portent sur des comparaisons entre une cohorte de patients atteints et une cohorte de patients sains, sans qu'il y ait nécessairement des relations de parenté entre les parents atteints. Les études d'association ont une meilleure puissance et une plus grande précision lorsque les polymorphismes que l'on cherche à détecter ont un risque relatif faible (c'est-à-dire un effet modeste sur le phénotype) ce qui est le cas dans les maladies complexes. Dans le cadre de l'approche gène-candidat, les études d'association consistent à comparer la distribution des allèles ou des génotypes d'un marqueur entre une collection de patients atteints de la maladie et une cohorte de témoins. L'augmentation statistiquement significative de la fréquence d'un allèle du marqueur indique que le

variant testé (méthode d'association directe) ou qu'un variant en déséquilibre de liaison avec le marqueur testé (méthode d'association indirecte) est un allèle de susceptibilité de la maladie. Comme les études d'association de gènes candidats utilisent quelques variants communs, au lieu de reséquencer des gènes entiers, elle est moins coûteuse que la méthode de reséquençage des gènes candidats. Néanmoins ces études d'associations de gènes-candidats nécessitent d'avoir prédit correctement l'identité du ou des gènes candidats, en général en se fondant sur des hypothèses physiopathologiques ou bien en utilisant des gènes candidats déterminés dans une précédente étude. Ce type d'études d'associations limitées à des gènes-candidats posent donc au moins trois types de problème (1) elles ne peuvent être mises en place que si l'on dispose de connaissances physiopathologiques préalables sur la maladie (2) même si elles aboutissent à identifier quelques mutations délétères sur les gènes-candidats, elles ne garantissent pas l'identification de tous les facteurs génétiques impliqués dans la maladie – seulement d'une fraction d'entre eux (3) enfin, ces études sont relativement difficiles à répliquer à cause de problèmes d'échantillonnages assez complexes sur lesquels nous ne nous attarderons pas ici mais qui complexifient encore l'analyse des résultats de ces études d'association génétique (Cordell et Clayton, 2005 ; Tabor *et al.,* 2002).

Annexe 2 : (Chapitre 4) Prémisses nécessaires pour accepter le concept de théorie médicale

Dans la communauté des philosophes de la médecine, il existe un scepticisme fortement répandu à l'égard de l'existence et/ou de l'utilité des théories médicales, ce qui explique peut-être que les auteurs qui se sont intéressés à la définition et à la structure des théories médicales soient peu nombreux (Lemoine, à venir). L'un des objectifs de cette thèse est de démontrer qu'il existe des théories médicales et qu'elles sont utiles. C'est pourquoi nous définissons au chapitre 4 notre concept de théorie médicale et nous montrons aux chapitres 5 et 6 qu'il existe au moins deux exemples de théorie médicale. Cette annexe peut donc sembler superflue. Il nous a cependant paru utile de renforcer notre démonstration, en nous attaquant aux sources de ce scepticisme protéiforme. Notre objectif est de montrer qu'accepter le concept de théorie médicale ne suppose pas d'embrasser des prémisses très fortes sur la structure du savoir médical ou sur l'analyse des concepts de santé et de maladie. Autrement dit, nous voulons démontrer que le concept de théorie médicale n'est pas nécessairement un concept clivant en philosophie de la médecine et peut être adopté par des philosophes de sensibilités très différentes.

En effet, il existe principalement trois versions du scepticisme à l'égard des théories médicales, selon que les auteurs abordent la question des théories médicales en questionnant a) le statut scientifique de la médecine, b) le problème de l'analyse des concepts de santé et de maladie, c) le problème de la démarcation et de la définition des maladies individuelles. En examinant les différents arguments des sceptiques, cette section nous permettra de dégager en creux les quelques présupposés qu'il est nécessaire d'accepter pour penser le concept de théorie médicale et qui sont étonnamment moins nombreux que ne pourrait le faire penser la popularité de la position sceptique. Ces présupposés sont au nombre de trois : le concept de théorie médicale a) suppose de reconnaître l'existence d'une science médicale qui ne soit pas entièrement subordonnée à la pratique clinique, b) est plus facilement accepté par les tenants d'une définition biologique du concept de maladie

mais peut être accepté en principe par quiconque accepte une distinction possible entre fait et valeur en médecine, c) peut être adopté, indépendamment de la question de savoir si les maladies individuelles sont des espèces naturelles.

1. Le concept de théorie médicale suppose de reconnaître que la constitution du savoir médical n'est pas entièrement subordonnée à la pratique clinique

Une des plus vieilles discussions en médecine concerne la distinction du statut de la médecine : est-ce une science ou un art ? La question peut paraître dépassée, dans la mesure où semble régner un consensus sur le statut hybride de la médecine. Ainsi, il est courant de distinguer d'une part la « science médicale », qui fait référence à la recherche biomédicale qui a lieu en laboratoire et plus généralement à l'idée d'une forme de connaissance « pure » des mécanismes physiopathologiques et de l'autre, la « pratique clinique », qui fait référence à l'activité du médecin à l'hôpital et dans son cabinet et plus généralement à des activités telles que le diagnostic, le pronostic et la thérapeutique. Cependant, cet apparent consensus cache en fait des différences subtiles, selon que les philosophes accordent la priorité à l'une de ces deux branches de la médecine et selon la façon dont ils perçoivent les relations entre ces deux branches. Et ces différences subtiles sont tout à fait déterminantes quant au débat sur l'existence des théories médicales. Accepter l'existence de théories biomédicales suppose en effet de reconnaître, sinon une priorité de la science médicale sur la médecine clinique, du moins une certaine forme d'indépendance entre ces deux domaines, afin d'affirmer que l'élaboration de la connaissance médicale n'est pas entièrement subordonnée aux objectifs et à la structure de la médecine clinique et qu'il existe un objet spécifique des théories médicales. Par opposition, les auteurs qui sont sceptiques vis-à-vis du concept de théorie médicale affirment en général une priorité de la pratique clinique sur la science médicale. Cette supposée priorité de la médecine clinique sur la science médicale consiste à affirmer que la médecine n'a pas pour objectif premier la connaissance scientifique des mécanismes physiopathologiques des maladies, mais seulement le traitement efficace des individus singuliers. C'est ainsi que Munson, qui va jusqu'à mettre en doute l'intérêt de financer une recherche fondamentale en médecine, écrit :

> « Il ne devrait pas être surprenant que la médecine contemporaine, en dépit de son engagement vis-à-vis de la science, continue à employer le succès en pratique ou le contrôle comme son critère fondamental pour évaluer les règles, les procédures et les affirmations causales. La médecine est une entreprise éminemment pratique qui doit s'efforcer de satisfaire des exigences immédiates et urgentes. Elle ne peut pas se permettre d'attendre l'acquisition d'une connaissance scientifique appropriée mais doit faire du mieux qu'elle peut lorsqu'elle est confrontée à l'ignorance et à l'incertitude. Nous n'attendons pas de la médecine qu'elle nous dise comment est le monde. Nous comptons plutôt sur la médecine pour agir contre la maladie et la souffrance ». (Munson, 1981, p. 195)

L'argument de Munson est double : d'une part, il insiste sur le fait que la finalité de la médecine n'est pas l'acquisition du savoir médical mais bien l'acquisition d'un art thérapeutique et d'autre part il souligne que dans de nombreux cas, la pratique clinique ne nécessite pas de reposer sur une connaissance scientifique ou sur des théories médicales pour être efficace. C'est ainsi qu'un certain nombre de médicaments sont utilisés sans qu'on connaisse exactement leur mécanisme d'action thérapeutique (qu'on pense à l'aspirine, par exemple, dont l'ensemble des mécanismes d'action est encore mal compris dans de nombreuses indications ou à l'usage de l'électrothérapie dans certains états de dépression sévère) – seul compte le soulagement thérapeutique. On peut enfin trouver un troisième argument qui s'appuie sur les deux premiers et renforce l'idée d'une prééminence de la pratique clinique sur la connaissance scientifique : la recherche médicale est elle-même dépendante des questions que pose la pratique clinique et ce faisant, elle s'appuie sur de nombreuses disciplines dont elle utilise les connaissances fondamentales pour la résolution de « problèmes cliniques » (Engelhardt, 1984) sans mener (ou de façon accidentelle) à la production de connaissances théoriques générales. C'est ce que Miller Brown exprime, en insistant sur le statut de « science appliquée » de la recherche médicale et en distinguant « les théories *en* médecine » du concept de « théorie médicale » :

> « La médecine est largement une activité pratique, puisqu'elle se préoccupe du diagnostic et du traitement de la maladie, le plus souvent dans un contexte de crise médicale ; et, en tant que tel, elle est plus proche de la technologie. La théorie *en* médecine, comme en électronique, emprunte aux sciences fondamentales comme la biologie, la chimie et la physique. La recherche en médecine, quand ce n'est pas de la biologie, de la chimie ou de la physique, est une sorte d'entreprise technologique alliée à ces sciences et mène rarement directement à un développement théorique… En tant que discipline pratique, la médecine et ses concepts de « maladie » et de « santé » sont intrinsèquement liés à la pratique médicale et aux intérêts des patients et des médecins, aussi bien qu'aux avancées scientifiques. Et c'est cela qui ajoute à la complexité et à la diversité qui font échec aux efforts pour trouver des définitions simples. » (Brown, 1985, p. 346)

Ces trois arguments ont été repris dans un article d'Harold Kincaid (Kincaid, 2008), qui va jusqu'à affirmer que, dans l'hypothèse même où il y aurait des théories médicales, celles-ci ne seraient pas utiles. Le point de départ de l'argumentation de Kincaid est une remise en question de la façon dont la philosophie de la médecine se serait alignée sur une philosophie générale des sciences historiquement fondée sur l'analyse de la physique. La médecine est une science appliquée, intéressée d'abord par la découverte de causes que l'on peut manipuler et sur lesquelles on peut intervenir, et qui s'attaque à un objet complexe : le vivant et ses chemins métaboliques complexes et multiples, faits de rétrocontrôle et de pléiotropie. Il serait donc absurde, alors même qu'on s'accorde à dire que la connaissance scientifique peut se passer de lois et de théories, d'exiger de la médecine qu'elle se conforme aux standards posés pour les sciences physiques. Ceci étant posé, adoptant une position descriptive, Kincaid choisit l'exemple du cancer pour montrer que :

1) Il n'y a pas de théorie du cancer et plus généralement, ce que la médecine appelle des théories n'en sont pas au sens classique du terme mais constitue seulement « des petits bouts d'explication causale » :

> « De la même façon, il n'y a pas de théorie simple de la maladie et du fonctionnement normal dont elle est la déviation. À la place, nous avons des petits bouts d'explications causales, et celles-ci sont souvent limitées à dire que tel gène ou tel produit du gène est « impliqué », sans beaucoup de détail sur *comment* il est impliqué. Fréquemment, nous ne pouvons pas dire au niveau moléculaire quelle est la cause du cancer et quel est son effet. Bien que nous puissions avec certitude identifier certaines cellules comme cancéreuses, nous n'avons pas d'explication générale du moment où le fonctionnement normal se termine et où le cancer commence. » (Kincaid, 2008, p. 373)

2) Mais que le fait de ne pas avoir de théorie du cancer, c'est-à-dire de ne pas disposer d'une description détaillée et cohérente des mécanismes généraux impliqués dans tous les cancers, n'a pas empêché la recherche médicale de développer des traitements qui sont efficaces contre la maladie :

> « Finalement, lorsque nous avons une connaissance causale solide, c'est souvent la connaissance que des traitements ont un certain effet, sans savoir exactement comment ils parviennent à cet effet. Environ 50% des traitements anticancéreux actuels sont des produits naturels dont on a dépisté systématiquement l'activité antitumorale sans avoir de bonne connaissance causale de comment ils fonctionnent, mais seulement la connaissance qu'ils fonctionnent. Ce type de cause qui peut être manipulé – un traitement – entraîne bien sûr la science avec elle à la recherche de marqueurs physiologiques prédictifs, et non pas dans une tentative de produire une compréhension théorique simple et élégante, quand bien même des chercheurs sur le cancer affirmeraient que c'est ce qu'ils produisent. » (Kincaid, 2008, p. 373)

On pourrait d'abord reprocher à Kincaid une double mauvaise foi. Il choisit en effet l'exemple du cancer, une maladie extrêmement hétérogène sur le plan moléculaire, biologique ou clinique, qui est presque davantage « une classe de maladie » qu'une maladie individuelle : trouver une théorie générale du cancer est

donc une tâche particulièrement complexe. Par ailleurs, s'il n'existe pas de théorie unique du cancer qui permette de subsumer toutes les instances de cancer sous une même explication, il a existé et il existe certainement plusieurs théories du cancer. Ainsi l'hypothèse de Knudson (Knudson, 1971) par exemple qui explique l'apparition du cancer en « deux coups » (deux mutations sont nécessaires, l'une qui active un proto-oncogène et l'autre pour désactiver un anti-oncogène) ou plus récemment, la théorie des cellules souches cancéreuses, sont des théories sur le cancer. Les différentes théories sur le cancer sont plus ou moins compatibles, parviennent à expliquer plus ou moins bien les différents types de cancer, mais qu'elles soient acceptées ou rejetées, elles ont contribué à une meilleure compréhension de la complexité des mécanismes à l'œuvre dans les cancers.

En dehors de ces arguties sur le choix de l'exemple du cancer, il est surprenant lorsqu'on examine la première étape de l'argumentation de Kincaïd, de constater qu'un auteur qui défend la spécificité de la philosophie de la médecine vis-à-vis de la philosophie des sciences, choisit de s'appuyer sur une conception syntaxique des théories scientifiques qui prend justement les théories physiques pour modèle :

> « Avons-nous ou pas une théorie du cancer, au sens où nous avons la biochimie et la biologie moléculaire ? J'aurai tendance à considérer que la biologie moléculaire et la biochimie n'ont pas grand chose à voir avec des théories pour commencer. En effet, elles nous fournissent une large palette de détails causaux spécifiques à propos de processus cellulaires particuliers et non un ensemble de principes généraux formalisables qui unifient les phénomènes. Bien sûr, plus nous nous rapprochons d'explications purement physiques et plus nous nous rapprochons d'une théorie générale. Cependant, ce n'est pas une théorie sur la façon dont la maladie fonctionne et sur comment cet état diffère du fonctionnement normal. La médecine peut *faire usage* des théories qui viennent d'ailleurs, mais ce n'est pas avoir une théorie *de la* médecine. » (Kincaid, 2008, p. 374)

C'est d'ailleurs un trait communément partagé par les tenants de cette position sceptique : tout se passe comme si le seul point de comparaison possible avec les

théories médicales était précisément la conception la plus exigeante des théories scientifiques, alors même que, comme on l'a rappelé plus haut, l'existence de théories biologiques est communément acceptée dans la communauté scientifique, à condition qu'on ne cherche pas à s'appuyer sur l'existence de lois en biologie et qu'on cherche ailleurs les fondements des explications d'un certain niveau de généralité qu'on peut trouver dans les sciences du vivant.

Cette deuxième étape de l'argumentation de Kincaid nous paraît encore plus discutable. En effet, admettre la finalité thérapeutique de la médecine ne prouve en rien que la recherche biomédicale ne puisse tirer profit de l'investigation des mécanismes physiopathologiques généraux des maladies. De la même manière, si nous reconnaissons l'efficacité de certains traitements empiriques, il est bon de rappeler que fonder un traitement sur la démonstration empirique de son efficacité n'est certainement pas sans risque et qu'au contraire, connaître les mécanismes physiopathologiques par lesquels un traitement est efficace permet de gagner aussi bien en efficacité et en précision thérapeutique qu'en fiabilité – un point essentiel justement au regard de la finalité thérapeutique de la médecine. Prenons ainsi le célèbre exemple des médicaments bétabloquants dont on a pensé dans les années 1990 qu'ils réduisaient la mortalité cardiovasculaire péri-opératoire avant de réaliser qu'ils augmentaient aussi considérablement le risque d'accident vasculaire cérébral en particulier à cause de leurs effets hémodynamiques qui avaient largement été sous-estimés.

Ainsi, ce qu'Harold Kincaid prouve, comme les autres tenants de cette forme de scepticisme – nonobstant qu'ils se gardent de mentionner l'apport de théories médicales comme la théorie des germes au développement de la prise en charge thérapeutique des maladies infectieuses au début du siècle –, c'est que a) s'il existe des théories médicales, celles-ci n'auront pas la forme de théories physiques du fait même de leur objet incroyablement complexe qu'est le vivant – une affirmation qui est parfaitement acceptée pour les théories biologiques b) que la médecine a pu développer des traitements efficaces en se fondant sur une connaissance causale partielle ou sur une évaluation empirique. Il n'y a ici aucun argument qui établisse l'impossibilité de principe des théories médicales ou qui établisse qu'une théorie

médicale ne puisse pas permettre de développer de nouvelles thérapeutiques efficaces.

En résumé, le concept de théorie médicale ne suppose ni d'affirmer que la finalité de la médecine consiste à établir un savoir médical, ni de minimiser la finalité pratique et donc thérapeutique de la médecine, ni de nier l'efficacité thérapeutique de traitements développés de façon totalement empirique. Elle suppose en revanche d'accepter que la constitution de ce savoir médical ne soit pas entièrement subordonnée à la finalité thérapeutique et que les théories médicales ne sauraient avoir la même forme que les théories physiques[64].

2. Le concept de théorie médicale suppose qu'on puisse distinguer – au moins provisoirement – les faits des valeurs dans une analyse des concepts de santé et de maladie

Peut-on formuler une théorie médicale sans définir le concept de maladie ? Le débat qui a majoritairement structuré le champ de la philosophie de la médecine ces trente dernières années concerne l'analyse des concepts de maladie et de santé[65]. Dans sa version la plus récente[66], c'est-à-dire à partir des années 1970, le débat s'est historiquement construit sur la méthode de l'analyse conceptuelle, qui consiste à proposer un ensemble de critères nécessaires et suffisants pour définir le concept de maladie, de façon à ce que la définition proposée corresponde à nos intuitions sur ce que nous appelons « maladie » et puisse englober l'ensemble des instances que nous considérons être des « maladies ». Maël Lemoine compare cette méthode à une forme de « jeu » dont il définit les règles comme une bataille entre définitions à partir

[64] Nous laissons ici provisoirement de côté la distinction entre théories « en » médecine et théories médicales, c'est-à-dire la discussion du caractère interdisciplinaire de la médecine, que nous abordons au moment de distinguer les théories médicales des théories biologiques dans la seconde section de ce chapitre.

[65] Pour une présentation synthétique mais précise de ce débat, on se reportera à l'analyse d'Elodie Giroux qui consacre une section entière d'un chapitre dédié à la philosophie de la médecine (Giroux, 2010a) ou à l'introduction du recueil de textes réunis par Elodie Giroux et Maël Lemoine sur les concepts de santé, maladie et pathologie (Giroux et Lemoine, 2012)

[66] Pour des raisons de concision, nous n'abordons pas ici la version du débat qui a dominé la première partie du 20e siècle entre les tenants du modèle « biomédical » et ceux du modèle « bio-psycho-social » mais les deux références citées ci-dessus montrent parfaitement le lien entre les deux formulations du débat.

d'exemples et de contre-exemples de maladies qui ne satisfont pas les critères de la définition de l'opposant (Lemoine, 2013b).

Au sein de cette méthode de l'analyse conceptuelle, on considère en général que deux grandes familles de pensées se sont constituées dans le monde anglo-saxon[67] : les « naturalistes » et les « normativistes » dont les représentants les plus célèbres sont respectivement Christopher Boorse (Boorse, 1975, 1976, 1977, 1987, 1997, 2011) et Jérôme Wakefield (Wakefield, 2000, 1992a, 1992b, 2007) , du côté des naturalistes et Engelhardt (Engelhardt, 1975, 1984, 1996) et Nordenfelt (Nordenfelt, 1995, 2001, 2007), du côté des normativistes. Les naturalistes défendraient ainsi l'existence d'un concept scientifique de maladie qui ne relèverait que de faits biologiques objectifs, tandis que les normativistes considéreraient que le concept de maladie est intrinsèquement normatif, c'est-à-dire, intrinsèquement dépendant de valeurs sociales ou culturelles.

Si l'on présente ainsi le débat, on pourrait intuitivement penser que le concept de théorie médicale sera plus facilement accepté par les tenants du naturalisme, qui défendent un concept biologique de la maladie et qui prétendent définir le concept théorique de maladie (associé au terme anglais *disease* – qu'on peut traduire par « pathologie ») alors que les défenseurs du normativisme, attachés au concept pratique de maladie (alors associée au terme *illness*), considéreraient les théories médicales comme la justification scientifique de la médicalisation d'états pathologiques définis comme tels d'abord en fonction de valeurs.

La distinction entre ces deux grandes familles de pensées est cependant moins évidente qu'il n'y paraît, ne serait-ce que parce que les représentants du naturalisme comme du normativisme eux-mêmes ont d'emblée relativisé cette distinction. Ainsi, Christopher Boorse ne nie pas l'existence d'un concept pratique normatif de maladie, même s'il donne la priorité au concept théorique. Il distingue d'ailleurs dans ses opposants un « normativisme fort » (Boorse, 1975, p. 51) d'un « normativisme faible » : alors que les tenants du normativisme fort considèrent que la démarcation entre maladie et santé relève entièrement d'un jugement de valeur, les partisans du

[67] La tension entre normativisme et normativisme apparaît déjà à bien des égards dans la pensée de Georges Canguilhem, même si les philosophes anglo-saxons y font peu référence. Pour une analyse détaillée des rapports entre la pensée de Georges Canguilhem et le débat anglo-saxon contemporain, on se reportera à l'ouvrage d'Elodie Giroux (Giroux, 2010b) .

normativisme faible considèrent que le concept de maladie est à la fois un jugement descriptif (laissant la porte ouverte à une description biologique du concept de maladie) et un jugement normatif. Quant à Jérôme Wakefield, il s'appuie sur une définition du trouble pathologique comme « dysfonctionnement préjudiciable » (Wakefield, 1992b), qu'il qualifie lui-même de position « hybride » dans ce débat, puisqu'il articule un jugement descriptif (la notion de dysfonctionnement étiologique) à un jugement de valeur (la notion de « préjudice » renvoie aux valeurs subjectives du malade vis-à-vis de sa maladie). La distinction est d'autant plus compliquée que la maladie mentale occupe un statut à part dans la définition du concept de maladie. Ainsi, s'il est difficile de trouver des tenants d'un « normativisme fort » pour les maladies physiques dans cette littérature, certains défendent un normativisme fort pour la maladie mentale. C'est le cas de Thomas Szasz qui considère que la maladie mentale est un pur mythe (Szasz, 1960), entièrement déterminé par nos normes sociales et culturelles, mais qui parvient à cette conclusion en partant d'une conception naturaliste de la maladie physique définie comme une lésion. Dans la mesure où le substrat anatomique de la lésion de la maladie mentale est très difficile à mettre en évidence, Szasz en conclut que ce que nous appelons « maladies mentales » est un pur produit de nos jugements de valeurs. Szasz est donc à la fois naturaliste fort du point de vue de la maladie physique et normativiste fort du point de vue de la maladie mentale.

C'est parce que la distinction entre « normativisme » et « naturalisme » est relativement difficile à déterminer de façon non ambiguë que certains auteurs suggèrent de l'abandonner et considèrent que « c'est la question de la primauté donnée au composant biologique sur le composant pratique dans la conception de la maladie » (Giroux et Lemoine, 2012, p. 15) qui structure véritablement le débat[68]. Si on adopte cette distinction entre des conceptions prioritairement biologiques et des conceptions prioritairement pratiques de la maladie et de la santé, on comprend bien que si le concept même de « théorie médicale » en tant qu'explication scientifique naturelle des maladies s'accorde mieux en première intention avec les conceptions

[68] Pour une analyse détaillée de toutes les terminologies employées et les positions adoptées dans cette littérature, on se reportera à l'analyse de Hofmann qui considère que la question « Qu'est-ce que la maladie ? » peut être analysée comme un ensemble de six sous-questions qui expliquent la complexité d'une démarcation franche entre normativistes et naturalistes (Hofmann, 2001) .

prioritairement biologique de la santé et de la maladie, ce concept peut en fait probablement être adopté par tous ceux qui acceptent de distinguer les faits et les valeurs dans une définition de la santé et de la maladie. En ce sens, la possibilité d'une théorie médicale n'exclut en principe aucune position prédéfinie sur l'analyse des concepts de santé et de maladie, exception faite des normativistes forts, dont la position est plus que minoritaire.

Ce point apparaît très clairement si on suit la double distinction proposée par Dominic Murphy à propos du concept de maladie mentale (Murphy, 2005) : l'opposition entre « objectivisme » (le concept de maladie relève d'un jugement de faits) et «constructivisme » (le concept de maladie relève d'un jugement de valeurs), et l'opposition entre « conservatisme » (nos intuitions sur la démarcation entre maladie et santé sont bonnes, toute définition de la maladie doit donc partir de nos intuitions ou du moins s'y conformer) et « révisionnisme » (nos intuitions sur la distinction entre maladie et santé sont susceptibles d'être fausses et peuvent être remises en question), aboutissant à une catégorisation en quatre positions des définitions de la maladie :

a) Le révisionniste-objectiviste considère que l'analyse scientifique est première dans la définition du concept de maladie et est susceptible de remettre en cause nos intuitions.

b) Le conservatiste-objectiviste affirme que l'analyse scientifique est utile pour comprendre la démarcation entre la santé et la maladie mais doit être restreinte au cadre de nos intuitions sur la distinction entre santé et maladie

c) Le révisionniste-constructiviste considère que nous construisons des théories physiopathologiques « sélectives », c'est à dire qui sont bien fondés sur des faits naturels mais qui dépendent du fait qu'on a choisi préalablement de classer certains individus comme malades.

d) Le conservatiste-constructiviste considère que notre physiopathologie n'est que la somme des comportements sociaux que nous jugeons déviants.

Si nous mentionnons cette analyse de Murphy, c'est d'abord parce qu'elle conforte l'analyse d'Hofmann, Giroux et Lemoine : la distinction entre normativisme et naturalisme est moins évidente et moins tranchée qu'elle n'y paraît. C'est par ailleurs parce que cette typologie vient renforcer notre thèse selon laquelle accepter le

concept de théorie médicale ne présuppose pas que l'on ait une position spécifique sur l'analyse des concepts de santé et de maladie. En revanche, selon la position que l'on a adoptée sur l'analyse des concepts de santé et de maladie, une théorie médicale ne sera pas acceptée de la même manière. C'est ainsi que la position révisionniste-objectionniste est la plus favorable aux théories médicales : les théories médicales sont alors considérées comme le fondement de notre définition des concepts de maladie et de santé et cette définition est susceptible de changer à mesure que les théories médicales changent. Au contraire, la position conservatiste-constructiviste (mais qui encore une fois n'est adoptée que rarement dans le cas des maladies psychiatriques et jamais pour les maladies physiques) est clairement la plus défavorable aux théories médicales – celles-ci n'étant considérées que comme des entreprises de justification de nos tentatives sociales de médicaliser certains comportements. Il faut noter cependant que la charge de la preuve pèse sur les constructivistes et non sur les objectivistes. En effet, à partir du moment où sont élaborées des théories médicales qui ont un pouvoir explicatif, le travail des objectivistes est un travail classique d'épistémologie des sciences : il s'agit de savoir à quel point ces théories expliquent les maladies et comment elles les expliquent. Le travail des constructivistes en revanche est de prouver soit que les théories médicales proposées sélectionnent des faits naturels de façon à justifier des normes sociales, soit que les faits naturels n'en sont pas et que les maladies ne sont bien définies que par des jugements de valeur.

L'opposition entre « conservatisme » et « révisionnisme » implique une forme de parti pris vis-à-vis de l'analyse conceptuelle. Pour les conservatistes, nos intuitions sur la démarcation entre santé et maladie sont premières et toute définition de la santé et/ou de la maladie doit impérativement respecter ces intuitions. Pour les révisionnistes, au contraire, ces intuitions peuvent être un point de départ mais elles sont susceptibles d'être remises en cause, par l'analyse scientifique pour les objectivistes et par l'analyse des normes sociales et culturelles pour les constructivistes. Mais que se passe-t-il lorsque l'on choisit une troisième voix entre normativisme et naturalisme, consistant à remettre en cause la méthode de l'analyse conceptuelle ? Est-ce que ceux qui essaient d'envisager la définition de la maladie en

dépassant le cadre de l'analyse conceptuelle et de l'opposition entre normativistes et naturalistes acceptent le concept de théorie médicale ?

En effet, la méthode de l'analyse conceptuelle comme méthode de résolution de la controverse entre naturalistes et normativistes ou entre conceptions biologiques et conceptions pratiques ou entre objectivistes et constructionnistes, est de plus en plus remise en cause dans la littérature contemporaine (Schwartz, 2007 ; Ereshefsky, 2009 ; De Vreese, à venir ; Lemoine, 2013a). C'est ainsi que Maël Lemoine rappelle que la méthode de l'analyse conceptuelle n'a pas permis de trancher entre les conceptions biologiques et pratiques de la maladie. Les normativistes comme les naturalistes et comme les tenants d'une conception hybride continuent de penser que chacun dispose d'une définition satisfaisante de la maladie et que la définition de leurs adversaires ne couvre pas l'extension des termes de santé et de maladie. Ce constat implique trois hypothèses : (1) soit l'un des deux partis a fait une erreur dans l'usage de l'analyse conceptuelle, (2) soit les deux se sont trompés mais l'analyse conceptuelle reste la méthode de référence pour trancher le débat, (3) soit enfin, aucune erreur n'a été faite si ce n'est celle de penser que l'analyse conceptuelle pouvait être une méthode de résolution de la définition de la santé et de la maladie. C'est cette dernière hypothèse qui est défendue par Maël Lemoine, au motif que lorsqu'il y a un désaccord entre deux définitions de la maladie, deux possibilités surgissent : soit elles sont en désaccord sur l'intention du terme, soit elles sont en désaccord sur son extension. Lorsqu'elles sont en désaccord sur son extension, l'analyse conceptuelle ne peut pas décider quelle définition est la bonne puisqu'elle présume par définition l'extension des termes qu'elle analyse. Si au contraire les définitions sont en désaccord sur l'intention du terme, l'analyse conceptuelle ne peut là encore pas trancher car il faudrait montrer en quoi l'une est meilleure que l'autre, étant donné que le seul critère recevable dans cette méthode est d'objecter par l'extension. En revanche, si l'analyse conceptuelle ne peut permettre de trancher, trente ans de débat philosophique à ce propos ont permis de mettre en lumière un certain nombre de points d'accord entre naturalistes et normativistes :

> « Tout d'abord, tout naturaliste reconnaîtrait que dans la science, de nombreux termes sont empruntés au langage vernaculaire et sont redéfinis par la stipulation scientifique. De plus, tout naturaliste reconnaîtrait aussi que le langage vernaculaire est souvent connoté axiologiquement (qu'on pense à la connotation pré-darwinienne de « l'homme » par opposition à la définition darwinienne de « l'homme »). Enfin, tout naturaliste serait prêt à admettre que certains termes scientifiques, qui réfèrent d'abord seulement à des faits naturels, deviennent *aussi* connotés axiologiquement (qu'on pense à la connotation actuelle de « radioactif », par exemple). En particulier, tout naturaliste reconnaîtrait à la fois que « la maladie » a historiquement référé à des états indésirables et était alors un terme normatif, et qu'à la lumière des conséquences de ce type d'état pour les humains, la « maladie » sera toujours normative au sens où « radioactive » l'est, c'est à dire, non pas à cause de ce en quoi elle consiste, mais en raison de ce que sont ses connotations et ses conséquences.
>
> De l'autre côté, tout normativiste reconnaîtrait que, bien que « l'or » a des propriétés normatives (le fait d'être précieux, etc.), il a aussi une définition cohérente et propre en tant que fait naturel, comme l'élément avec le numéro atomique 79 par exemple. Tout normativiste reconnaîtrait aussi que lorsqu'on en vient à connaître les mécanismes d'une maladie et que le label correspondant devient une entité cohérente, une définition scientifique de la maladie comme fait naturel s'en suit. Cependant, certains normativistes continueraient d'affirmer que tout fait naturel est toujours une construction sociale, que « l'or » signifie essentiellement un métal précieux spécifique et signifie accidentellement le numéro atomique 79. Au minimum, ces constructivistes doivent reconnaître qu'il y a une différence entre deux types de présumés « faits construits » - ceux qui fonctionnent dans les explications scientifiques, les prédictions et les lois et les autres. Seule la première catégorie est appelée « fait naturel » par les naturalistes. » (Lemoine, 2013b, p. 321–322)

C'est de ce consensus sur le fait que toute maladie est porteuse de connotations normatives et qu'un nombre non négligeable de maladies ont été

« naturalisées » par une explication scientifique, que partent tous les tenants d'une troisième voix, entre naturalisme et normativisme qui cherchent à dépasser le cadre de l'analyse conceptuelle, qu'ils s'orientent plutôt vers une naturalisation de la maladie, comme Maël Lemoine (Lemoine, 2013a) ou vers une redéfinition pragmatique de la maladie, comme Leen De Vreese (De Vreese, à venir) .

Autrement dit, tous ceux qui s'essaient à définir les concepts de maladie et de santé, qu'ils choisissent de se placer dans le cadre du débat entre normativistes et naturalistes, ou qu'ils s'essaient à d'autres méthodes que celle de l'analyse conceptuelle, s'accordent sur une distinction entre faits et valeurs et sur la possibilité de distinguer dans nos définitions des maladies ce qui relève, comme le dit Ereshefsky de la description d'états d'une part et d'affirmations normatives d'autre part (Ereshefsky, 2009)[69]. Il suffit d'accepter cette distinction pour pouvoir accepter l'existence de théories médicales – quelle que soit la valeur ou le rôle qu'on donne ensuite à ces théories dans nos définitions de la santé et de la maladie.

3. Le concept de théorie médicale ne présuppose pas qu'on définisse les maladies comme des espèces naturelles

C'est un des débats récurrents des trente dernières années[70] que de savoir à la fois si *les maladies* sont des espèces naturelles ou pas. Précisons pourquoi ce point est important pour notre propos. Indépendamment de la façon dont on caractérise une « espèce naturelle » (existence d'une structure explicative sous-jacente, objet doté de propriétés naturelles, objet dont l'explication fait référence à des lois, etc.) – un point qui fait particulièrement débat et qui complique la question générale de savoir si les

[69] Il faut mentionner ici qu'Ereshefsky fait une distinction entre la dichotomie faits/valeurs et la dichotomie descriptions d'états/affirmations normatives au motif que de nombreuses descriptions d'états contiennent des affirmations normatives explicites : la distinction entre descriptions d'états et affirmations normatives se fait donc sur le caractère explicite du recours à des composants normatifs (Ereshefsky, 2009, p. 227).

[70] Cette question a déclenché les passions autour de la controverse entre Reznek (Reznek, 1987, 1995) et D'Amico (D'Amico, 1995) . Nous laissons ici de côté la question plus spécifique de savoir si les maladies mentales sont des entités naturelles, une question déjà soulevée par Thomas Szasz (Szasz, 1968), devenue très populaire dans le champ de la philosophie de la psychiatrie mais qui soulève des problèmes spécifiques dans lesquels il n'est pas utile d'entrer ici. Nous utiliserons cependant certains des arguments utilisés par les philosophes de la psychiatrie et qui sont transposables dans le débat général de savoir si les maladies individuelles sont des entités naturelles ou pas.

maladies ou la maladie sont des espèces naturelles, on considère généralement qu'identifier un objet comme « espèce naturelle » consiste à légitimer la recherche d'une explication de cet objet et l'élaboration d'une théorie scientifique. Pour donner un exemple simple et non controversé, on considère en général que l'or est une espèce naturelle alors que les chaises ne sont pas des espèces naturelles – ce qui aboutit à l'affirmation qu'il y a une théorie scientifique de l'or alors qu'il est peu probable qu'il y ait une théorie des chaises :

> « L'or est une espèce naturelle mais les chaises ne le sont pas. Laissez-moi expliquer pourquoi. Une espèce naturelle est une classe d'objets qui partage une nature sous-jacente commune profonde ou théoriquement intéressante qui explique l'ensemble des propriétés que chaque membre de cette classe partage. Chaque pièce d'or est jaune, brillante, conduit l'électricité, est malléable, et ainsi de suite, et ces propriétés sont expliquées par le fait que l'or a une nature sous-jacente – un nombre atomique avec 79 protons dans son noyau. L'or des fous n'est pas de l'or parce qu'il n'a pas cette structure – c'est seulement de la pyrite. Les chaises, de l'autre côté, partagent une et une seule propriété – elles ont une structure qui leur permet de supporter un être humain en position assise. Chaque chaise a un ensemble de propriétés différentes, certaines sont faites en bois, d'autres en métal, ou en plastique, certaines sont combustibles, d'autres pas, certaines fondent, d'autres pas et ainsi de suite. A cause de ces différences, une théorie scientifique à propos des chaises n'aboutirait pas à grand chose – il n'y a pas d'ensemble de propriétés partagées qui suggèrent une forme de nature sous-jacente que toutes les chaises auraient en commun et qui expliquerait pourquoi cet ensemble de propriétés existe. Sans cet ensemble, il est peu probable que les chaises aient une quelconque structure sous-jacente commune. » (Reznek, 1995, p. 571–572)

C'est précisément ce lien entre « espèces naturelles » et théorie scientifique qui nous préoccupe : faut-il considérer que les maladies sont des espèces naturelles pour accepter l'idée de théorie médicale et si on considère que les maladies ne sont

pas des espèces naturelles, est-on forcément opposé à l'idée de théorie médicale ? Le débat est souvent résumé à une opposition entre essentialisme et nominalisme. Les essentialistes considéreraient que les maladies sont des espèces naturelles, sur le modèle des éléments chimiques et que l'objectif de la médecine est de révéler les propriétés essentielles de ces espèces naturelles. Au contraire, les nominalistes (aussi appelés conventionnalistes) considéreraient que ce que les médecins appellent « maladies » relèvent de décisions socialement construites, en fonction de la délimitation des spécialités médicales, en fonction des conséquences préjudiciables pour l'individu atteint de certains symptômes ou en fonction des possibilités thérapeutiques. En pratique, la diversité des positions assumées aujourd'hui sur le problème des maladies comme espèces naturelles est bien plus vaste et peut être présentée ainsi : (1) La première position consiste à dire que les conceptions de ce qu'est une « espèce naturelle » se sont multipliées au point que plus personne ne sait ce qu'on entend par ce terme et qu'il ajoute plus de confusion que de clarté dans les débats où il est utilisé (Hacking, 2007). Une autre version de cette position, plus spécifique à la question des maladies, consiste à considérer que le concept d' « espèces naturelles » est un concept qui appartient à la philosophie générale des sciences, que l'on tente arbitrairement de plaquer en philosophie de la médecine (Kincaid, 2008). (2) La seconde position consiste à accepter le débat sur les espèces naturelles dans le champ de la philosophie de la médecine et à soutenir l'une des positions suivantes : 2a) ni les maladies individuelles, ni la maladie en générale, ne sont des espèces naturelles 2b) les maladies individuelles sont des espèces naturelles mais la maladie en général n'en est pas une (Reznek, 1987, 1995) 2c) les maladies individuelles et la maladie en général sont des espèces naturelles 2d) les maladies individuelles et la maladie en général ne sont pas des espèces naturelles mais sont définies par rapport à des espèces naturelles (Sulmasy, 2005). Les trois positions qui nous intéressent et que nous examinerons ici sont les positions 2a) et 2b) et 2c), car elles ont un impact direct sur le concept de théorie médicale. La position 2c) qui serait celle des naturalistes les plus convaincus impliquerait qu'on puisse avoir une théorie de la maladie, c'est-à-dire une théorie qui s'applique au concept de maladie en général et des théories de maladies individuelles (c'est-à-dire une théorie de la tuberculose, une théorie de la rougeole, etc.). La position 2b) qui est celle de Laurie Reznek par

exemple, impliquerait qu'on puisse potentiellement avoir une théorie des maladies individuelles mais pas de théorie générale de la maladie. Enfin, la position 2a) serait la plus radicale et consisterait à dire qu'il ne pourrait y avoir de théories en médecine, ni des maladies individuelles ni de la maladie en général.

Il est facile de remarquer le parallèle entre le débat essentialisme/nominalisme et le débat naturalisme/normativisme. Ce parallèle n'a rien d'hasardeux, puisque l'affirmation selon laquelle le concept de maladie constituerait une entité naturelle est utilisée comme un argument en faveur du débat normativisme/naturalisme[71]. Or, de la même manière qu'il est difficile de trouver des positions absolument naturalistes ou absolument normativistes, il est difficile de trouver des positions absolument essentialistes ou absolument conventionnalistes. Ainsi, de la même manière que notre objectif n'était pas de prendre position dans le débat naturaliste/normativiste mais de montrer que les deux positions partageaient au moins une position commune suffisant à accepter le concept de théorie médicale, notre argumentation consistera ici non pas à prendre position dans le débat essentialiste/nominaliste, mais à montrer quels sont les points de consensus du débat et en quel sens le concept de théorie médicale est indépendant du débat essentialiste/nominaliste. Nous procéderons en trois points : (1) Nous montrerons d'abord qu'il paraît impossible de démontrer a priori que ni les maladies ni la maladie ne sont des espèces naturelles, et qu'il est donc impossible de démontrer a priori qu'on ne peut pas avoir de théories médicales. (2) Dans un deuxième temps, nous présenterons un point du consensus actuel entre les essentialistes et les nominalistes : tous reconnaissent que, quel que soit le statut véritable des maladies (espèce naturelle, autre type d'entité, etc.), elles ont souvent été considérées comme des espèces naturelles et que ce rôle d'espèce naturelle a une dimension heuristique forte. Autrement dit, quand bien même les maladies ne sont pas des espèces naturelles, on ne perdrait rien à les considérer comme telles et ce d'autant plus que même les naturalistes les plus convaincus reconnaissent en même temps que nos classifications ne reflètent que très partiellement ces avancées scientifiques et qu'indépendamment de la question de savoir si les maladies sont des

[71] C'est à vrai dire le point de départ de la controverse entre D'Amico et Reznek : Reznek soutient que la maladie en général ne constitue pas une entité naturelle et utilise cette thèse pour affirmer la primauté de l'approche normativiste (Reznek, 1995) .

espèces naturelles ou pas, il y a une place pour des classifications pragmatiques. (3) Enfin, je montrerai que si considérer les maladies individuelles comme des espèces naturelles est certainement une incitation à développer des théories médicales, l'inverse n'est pas vrai : même lorsqu'on refuse de considérer les maladies individuelles comme des espèces naturelles, on peut défendre la recherche d'une explication scientifique des maladies. C'est par exemple la position de ceux qui défendent que les maladies sont des « espèces pratiques ».

(1) Il est impossible *a priori* de montrer que ni les maladies ni la maladie ne sont pas des entités naturelles.

Pour montrer qu'une maladie est une espèce naturelle, quelle que soit la façon dont on définit les espèces naturelles, il faut pouvoir démontrer que la définition de la maladie en question s'appuie sur un certain nombre de propriétés (ensemble de symptômes, causes, mécanismes biologiques) que toutes les instances de la maladie en question ont en commun. Selon la définition que l'on choisit des espèces naturelles, il est donc possible de prouver qu'une maladie est une espèce naturelle, lorsque l'explication scientifique de la maladie est suffisamment bien connue. Par exemple, Jonathan Tsou cherche à prouver que la dépression et le suicide sont des espèces naturelles (Tsou, 2013). Il définit les espèces naturelles comme *« des classes d'objets, de propriétés ou de processus qui arrivent naturellement et qui existent indépendamment des classificateurs qui les découvrent (par exemple, les électrons, l'or, l'eau, les poissons) »* (Tsou, 2013, p. 461), par opposition aux espèces « artificielles » qui sont des classes d'objets qui n'existent pas dans la nature et sont inventés par les classificateurs (comme les chaises). Pour prouver que la dépression et le suicide sont bien des espèces naturelles, il considère un ensemble de propriétés biologiques que l'on retrouve de façon stable dans les cas de dépression et de suicide et conclut que la dépression et le suicide sont bien des espèces naturelles puisque *« elles sont des classes de comportement anormal associé à une dysfonction de la sérotonine dans des structures diverses du cerveau (c'est-à-dire, les amygdales, le cortex cingulaire antérieur le cortex préfrontal et l'hippocampe) »* (Tsou, 2013, p. 461). On peut ou pas être d'accord avec Tsou sur la façon dont il conceptualise les espèces naturelles, ou sur

la façon dont il identifie des propriétés biologiques stables pour la dépression et le suicide, nous ne cherchons ici qu'à mettre en exergue le déroulé de son argumentation. En revanche, comment démontrer a priori qu'une maladie n'est pas une espèce naturelle ? Si on choisit une argumentation similaire, en identifiant d'abord un concept d'espèce naturelle puis en montrant que nos explications scientifiques ne permettent pas de mettre en exergue pour la maladie donnée le type de propriétés ou de caractéristiques qu'on a identifié dans le concept d'espèce naturelle, on n'aura rien prouvé ou plutôt on aura seulement prouvé qu'*aujourd'hui* telle maladie n'est pas une espèce naturelle. Mais on n'aura pas prouvé que la maladie en question ne pourra jamais devenir une espèce naturelle et on ne peut pas prouver qu'il n'existera jamais d'explication scientifique satisfaisante pour la maladie en question.

Le même problème se pose pour la question de savoir si la maladie est une espèce naturelle. Pour prouver que la maladie n'est pas une espèce naturelle, Reznek (qui a une position un peu ambiguë sur les maladies individuelles dont il concède que certaines sont des espèces naturelles mais que d'autres ne le sont pas) insiste sur le fait que lorsqu'on considère la diversité des maladies individuelles, on ne peut pas espérer trouver une structure sous-jacente commune à toutes les maladies individuelles et qui ferait du concept de maladie une espèce naturelle :

> « De plus, si nous regardons les natures de tous ces états que nous considérons être des maladies – par exemple, la schizophrénie, le paludisme, la syphilis, la sclérose en plaque, les maladies des neurones moteurs, la phénylcétonurie, l'épilepsie, les diabètes insulinodépendants, l'achondroplasie, et ainsi de suite – elles semblent si différentes qu'il est difficile de voir en vertu de quelle propriété partagée, leurs natures sont des maladies. » (Reznek, 1987, p. 70)

Or, comme le fait remarquer D'Amico, en mettant en avant la diversité des maladies individuelles, Reznek ne prouve pas de façon certaine que la maladie ne sera jamais une espèce naturelle et qu'il ne peut y avoir de théorie de la maladie, il prouve seulement qu'une telle théorie n'a pas encore été trouvée :

> « Mais les raisons de ses doutes [=les doutes de Reznek] sont l'échec d'une telle entité à correspondre aux classifications existantes, la diversité au niveau macroscopique de cette entité telle qu'elle est classée en termes de symptômes et le fait que toutes ces instances de classification sont de toute façon dépendantes de nos intérêts spécifiques (plus sur ce point en dessous). Ces raisons, cependant, ne peuvent établir à l'avance que l'investigation théorique de la maladie est dépourvue d'intérêt au sens où une étude théorique de la classification des objets « meubles » apparaîtrait dépourvue d'intérêt. Bien sûr, une science des meubles est dépourvue d'intérêt pour ces mêmes raisons que Reznek cite quand il discute les maladies, c'est-à-dire le fait que la diversité des classifications des meubles échoue à révéler une structure sous-jacente pour ces classifications. Mais dans le cas des meubles, il est à peine nécessaire d'argumenter ; nous avons précisément défini le concept de cette façon. Malgré tous les efforts de Reznek pour établir une intuition similaire à propos des maladies, il est nécessaire d'avoir une sorte d'argument conceptuel qui montre que la classification des maladies est exactement comme la classification de ces objets que sont les meubles, qui montre pourquoi ces concepts n'admettent pas d'espèces naturelles, et qui prouve ce point indépendamment des doutes que peuvent soulever les classifications existantes ou indépendamment du fait qu'une structure sous-jacente commune n'a pas encore été découverte ». (D'Amico, 1995, p. 560)

Dans sa réponse aux critiques de D'Amico, Reznek minimise ce point, arguant que la diversité des maladies individuelles prouve au moins avec une grande probabilité qu'il n'y a pas de théorie de la maladie possible et que nous avons donc de bonnes raisons empiriques de ne penser que nous n'aurons pas de théorie de la maladie (Reznek, 1995, p. 577). Il nous semble pourtant que l'argument de D'Amico est essentiel : étant donné la complexité et la diversité des maladies, l'absence de structure explicative commune évidente ne suffit pas à déterminer une fois pour toutes que les maladies ou la maladie ne sont pas des espèces naturelles et qu'on ne peut avoir de théories médicales. C'est d'autant plus vrai que la majorité des auteurs

reconnaissent que les maladies ont été considérées comme des espèces naturelles et que ce rôle historique a été déterminant dans certaines découvertes médicales d'importance comme dans la mise en place de la théorie des germes.

(2) Les maladies individuelles ont joué le rôle heuristique d'entité naturelle dans l'histoire de la médecine

Un des arguments les plus souvent utilisés pour affirmer que les maladies ne sont pas des espèces naturelles est l'évolution de notre classification des maladies – c'est ce que nous appelons l'objection des nosologues pragmatiques. Il est généralement considéré que le dix-neuvième siècle avec les postulats de Koch et la naissance de la théorie des germes a permis l'émergence d' « un modèle monocausal » (Broadbent, 2009 ; Carter, 2003) guidé par ce que Severinsen (Severinsen, 2001) appelle « le principe du mécanisme de la maladie », que Carter appelle « l'hypothèse de distinguabilité » (Carter, 2003), et qui consiste à affirmer qu'il existe une et au moins une seule cause pour une maladie donnée et par conséquent à distinguer deux maladies si elles ont deux causes différentes. Ce modèle a permis de grandes avancées scientifiques mais ne peut plus être aujourd'hui utilisé que ce soit pour décrire nos classifications ou pour normer la façon dont il faut classer les maladies. En effet, le modèle monocausal qui a prévalu au moment de la mise en place de la théorie des germes est peu opérant quand on est confronté à la multifactorialité qui caractérise la plupart des maladies chroniques non communicables telles que les cancers, les maladies cardiovasculaires, etc. Si on ne peut définir les maladies par une cause unique ou par un mécanisme biologique stable, Severinsen suggère de développer des classifications pragmatiques qui prennent en compte la démarche diagnostique et thérapeutique en priorité. Ces classifications pragmatiques sont susceptibles d'évoluer en fonction des intérêts des spécialités médicales, des systèmes de soin et des intérêts des différents acteurs des systèmes de santé : en définissant les maladies sur ces critères, les nosologues pragmatiques prouveraient que les maladies ne sont pas des entités naturelles, mais plutôt des processus nommés par convention et pour permettre aux différents acteurs des systèmes de santé de communiquer entre eux.

Les essentialistes ne rejettent pas ces arguments mais ils considèrent au contraire que l'évolution de nos classifications est au contraire la preuve que nos définitions des maladies ne sont pas seulement le reflet de considérations pragmatiques, mais aussi le reflet de nos avancées scientifiques, ce qui serait une caractéristique des espèces naturelles :

> « Deuxièmement, l'histoire de la médecine semble appuyer la vision naturaliste et non la vision conventionnaliste. À la lumière de recherches postérieures, les classifications passées ont révélé avoir groupé de façon erronée plusieurs groupes de maladies en une sur la base d'observations grossières récurrentes (comme dans le cas de la syphilis et de la gonorrhée), avoir identifié de façon erronée des symptômes avec des maladies (comme dans le cas des « fièvres ») et avoir manqué d'un traitement efficace jusqu'à l'identification de la maladie (comme pour le traitement de la tuberculose). Il n'est pas inhabituel dans l'histoire de la médecine que des états soient enlevés de la catégorie des maladies (comme la conversion hystérique) ou placés dans cette catégorie après n'avoir longtemps pas été considéré comme des maladies (l'épilepsie). Ces événements dans l'histoire de la médecine appuient fortement l'affirmation selon laquelle l'identification d'une maladie n'est pas une décision purement conventionnelle, dans laquelle une réforme de l'usage n'aurait pas d'autre intérêt que des raisons de convenance ou des raisons pratiques. Les divers systèmes de classification dans l'histoire de la maladie ne viennent pas, dans la plupart des cas, de pure convenance ou d'un choix arbitraire, mais s'appuient d'abord sur des descriptions proposées avant qu'on accède aux processus causaux sous-jacents de ces états. Une fois que la nature sous-jacente est découverte, idéalement un virus spécifique ou une bactérie, la description antérieure est soit réformée soit ajustée au nouveau processus pathologique maintenant découvert. » (D'Amico, 1995, p. 556–557)

Il faut bien noter que D'Amico ne nie pas que nos classifications médicales sont en partie dépendantes de descriptions symptomatiques (une maladie est définie phénoménologiquement en fonction d'un regroupement de symptômes partagés par

l'ensemble des patients) ou de choix pragmatiques. L'opposition de D'Amico tient à la signification de cette dimension pragmatique :

> « Il y a aussi des conventions dans les sciences naturelles qui concernent, par exemple, les unités de mesures pour lesquelles on peut dire que les théories « imposent » ou « inventent » une échelle de mesure pour le monde. Reznek affirme que l'idée d'espèce naturelle doit exclure une telle diversité dans la façon dont le monde est conceptualisé et toute décision conventionnelle dans la description théorique de la nature d'une chose. Mais tout ce que l'argument naturaliste requiert, quelle que soit l'approche théorique qui revaut, et quels que soient les intérêts humains qui jouent un rôle dans une enquête de ce type, c'est que la nature sous-jacente de la chose, s'il y en a une, ne soit ni inventée, ni imposée. Le naturaliste n'a pas besoin de défendre l'idée selon laquelle il n'y a qu'une seule façon de produire une description du monde naturel ou selon laquelle les classifications d'objets naturels doivent exclure tous les intérêts humains. » (D'Amico, 1995, p. 565)

Pour les essentialistes, donc, l'objection des nosologues pragmatiques n'en est pas une. Au contraire, l'histoire de la médecine montre que les maladies ont été considérées comme des espèces naturelles et que le fait de considérer les maladies comme des espèces naturelles, en appliquant par exemple, le modèle monocausaul et le principe de distinguabilité a eu un rôle heuristique considérable :

> « Apparemment, donc, un but de la médecine a été d'identifier les maladies réelles. De la même manière que deux échantillons géologiques tendent à partager un ensemble de propriétés parce qu'ils sont des échantillons du même minéral, deux patients tendent à partager certaines propriétés parce qu'ils ont la maladie. La réponse d'un patient à une thérapie donnée est (toutes choses égales par ailleurs) est pertinent du point de vue de la confirmation pour prédire la réponse d'un patient, du fait que les deux patients ont la même maladie. Les maladies ont fonctionné comme des entités naturelles dans le raisonnement médical. » (Lange, 2007, p. 277)

On peut donc résumer ici le point de consensus qui unit les positions essentialistes et nominalistes : a) nos classifications des maladies sont en partie dépendantes de critères pragmatiques b) les maladies ont joué le rôle heuristique d'espèces naturelles dans l'histoire de la maladie. Ce point est pour nous décisif : après avoir montré en (1) qu'il était difficile de montrer *a priori* que ni les maladies ni la maladie n'étaient des espèces naturelles et donc de démontrer a priori l'inutilité des tentatives d'explication des maladies et de l'enquête scientifique, nous montrons (2) qu'on ne perd rien à considérer les maladies comme des espèces naturelles. Nous allons montrer à présent qu'on peut adopter une position non-essentialiste et accepter le concept de théorie médicale (3).

(3) On peut ne pas être essentialiste et défendre le concept de théorie médicale

Le débat entre essentialisme et nominalisme est en partie structuré autour de l'affirmation selon laquelle si on considère qu'un objet est une espèce naturelle, alors il est légitime de chercher une théorie scientifique de l'objet en question. L'inverse n'est cependant pas vrai : si on ne considère pas un objet comme une espèce naturelle, on peut néanmoins considérer comme légitime de rechercher une théorie de cet objet. Autrement dit, dans notre discussion, adopter une position non-essentialiste à l'égard des maladies ne signifie pas que l'on nie l'intérêt des théories médicales. Nous allons discuter deux positions non-essentialistes ici : celle proposant de considérer les maladies psychiatriques comme des « entités pratiques » (Zachar, 2000, 2002 ; Zachar et Kendler, 2007) et celle qui propose de considérer les maladies psychiatriques comme des entités qui ont un ensemble de propriétés mécaniques (Kendler *et al.,* 2011).

Ainsi, en définissant les maladies psychiatriques comme des « espèces pratiques » (*practical kinds*), Zachar affirme non seulement qu'il ne s'oppose pas à l'explication scientifique des maladies psychiatriques, mais encore que cette définition rendrait davantage justice aux avancées scientifiques que la définition des maladies psychiatriques comme espèces naturelles. La définition des maladies psychiatriques

comme des espèces pratiques ne reviendrait pas à nier toute structure inhérente et stable aux maladies psychiatriques comme le ferait une position purement nominaliste, mais à laisser de côté toute affirmation métaphysique quant à la nature réelle ou pas des maladies diagnostiquées et à se concentrer sur la capacité des catégories psychiatriques à répondre au mieux aux intérêts pragmatiques de nos classifications médicales. L'argument principal de Zachar est de dire que les définitions essentialistes de la maladie fixent les contours des maladies une fois pour toutes, ce qui est peu compatible avec a) le fait que les maladies sont en partie le produit de l'évolution et sont donc amenées à changer à travers le temps b) le fait que nos théories scientifiques des maladies psychiatriques évoluent avec le temps et en fonction des décisions sociales qui nous amènent à considérer que tel état pathologique est une maladie ou pas :

> "Ni le modèle médical traditionnel et ses méthodes pour isoler la maladie ni l'approche mathématique psychologique des classifications ne sont parvenus à isoler ce que nous appellerions des *espèces naturelles.* Mais les deux ont découvert des comportements stables qui sont plus que de simples inventions, mais l'idée d'une réalité inhérente isolée, définissable uniquement en utilisant l'expérimentation biologique et des analyses statistiques sophistiquées est erronée. Nous avons besoin de trop d'autres variables et de formes de preuve pour individuer ces comportements et adopter différentes méthodes ou prioritariser certains types de preuve peut altérer les comportements que nous trouvons. Meehl (1986) a discuté les avantages de penser les taxons diagnostiques comme des concepts ouverts et a considéré que toute autre stratégie serait "scientifiquement douteuse". La réalité sera toujours plus complexe que ce que nos catégories capturent." (Zachar, 2000, p. 180)

Ainsi, conceptualiser les maladies comme des espèces pratiques permettrait de conceptualiser nos catégories comme « ouvertes » au changement scientifique et à la multiplicité des théories proposées pour expliquer les maladies psychiatriques. Une des difficultés cependant que pose le modèle des espèces pratiques est de ne pas

guider les principes qui devraient guider nos classifications. C'est pourquoi Zachar, Kendler et Craver ont proposé de considérer les maladies comme des espèces qui ont un ensemble de propriétés mécaniques (MPC kinds – *Mechanistic Property Cluster kinds*), une définition qui met particulièrement en valeur l'idée de structures explicatives stables des maladies :

> « Le changement que nous recommandons pour la psychiatrie, inspiré par le philosophe Richard Boyd (Boyd 1991, 1999 ; Wilson *et al.,* 2007) est de passer d'une quête des essences des entités psychiatriques, qui s'appuie sur des faits biologiques ou sociaux à propos de ces maladies, à une quête pour les mécanismes causaux complexes et multi-niveaux qui produisent, sous-tendent et maintiennent les syndromes psychiatriques (Kendler, 2008). […] Les ensembles de propriétés n'ont pas des essences simples et déterministes. Il n'est pas nécessaire que tous les membres recoupent un seul ensemble de traits ; les membres seraient plutôt regroupés les uns à côté des autres dans un espace objet à causes des mécanismes causaux et des contraintes du développement, de l'évolution et de la physiologie. Dans les limites de la simplicité et de la détermination, les MPCs tendent vers des essences, avec des propriétés et des mécanismes communs à l'ensemble des membres de l'espèce. À l'autre extrême, les entités tendent vers des espèces pratiques ou construites, où les limites entre les catégories sont souvent définies en fonction des pratiques classificatoires de certains agents. La vision MPC encourage l'idée qu'il y a des structures explicatives robustes à découvrir qui sous-tendent la plupart des maladies psychiatriques. Cependant, elle nous met en garde et nous prévient que ces structures seront désordonnées et qu'il faudra beaucoup de travail et un certain degré d'idéalisation et d'abstraction pour les mettre en lumière. » (Kendler *et al.,* 2011, p. 1146)

Pour résumer, si certains auteurs utilisent le débat sur le statut des maladies comme espèces naturelles pour développer un certain scepticisme vis à vis des théories médicales, cet argument nous paraît infondé pour trois raisons. D'abord, il est impossible de montrer *a priori* qu'une maladie, que nous ne considérerions pas

comme une espèce naturelle aujourd'hui, n'est pas susceptible de l'être en fonction de nos avancées scientifiques futures. D'autre part, il est existe une forme de consensus sur le fait que les maladies ont été considérées pendant longtemps comme des espèces naturelles et que ce statut a joué un rôle heuristique déterminant dans l'histoire de la médecine. Enfin, il est possible de considérer que les maladies ne sont pas des espèces naturelles et de défendre la nécessité de rechercher des structures explicatives stables des maladies.

Nous avons montré que le concept de théorie médicale, qui fait pourtant l'objet de diverses formes de scepticisme, ne s'appuie que sur un petit nombre de prémisses simples sur lesquelles de nombreux philosophes de la médecine peuvent s'accorder : pour accepter son existence et son utilité potentielles, il faut au minimum reconnaître que la science médicale n'est pas entièrement subordonnée à la pratique clinique, que l'on peut distinguer dans nos énoncés sur les maladies – au moins provisoirement - les faits naturels des valeurs, et que la façon dont on définit les maladies comme entités nosologiques est indépendante de la question des théories médicales, pourvu qu'on accepte que nos connaissances scientifiques peuvent modifier nos entités nosologiques et que celles-ci ne sont pas gravées dans le marbre une fois pour toutes mais sont amenées à évoluer, en partie du fait de l'évolution de nos connaissances scientifiques.

Glossaire

Pour formuler les définitions de ce glossaire, nous nous sommes inspirée de plusieurs ouvrages de biologie cellulaire (Alberts *et al., 2013 ;* Bassaglia, 2013). La plupart des définitions (sauf si le contraire est précisé) sont adaptées au cas des eucaryotes.

ADN : Acide désoxyribonucléique : molécule hélicoïdale double brin constituée d'un assemblage par paires de nucléotides. Un nucléotide est composé d'une association entre un sucre et un phosphate, liée à une base. On distingue quatre types de nucléotides notés A, T, C et G, du nom des bases correspondantes (l'adénosine, thymidine, cytosine et la guanine guanine). Chaque brin d'ADN possède une extrémité 5' phosphate et une extrémité 3' hydroxyle libre. Les orientations des deux brins de l'hélice sont en sens opposé : ils sont antiparallèles. Le génome humain comporte 6 milliards de bases réparties en 46 chromosomes.

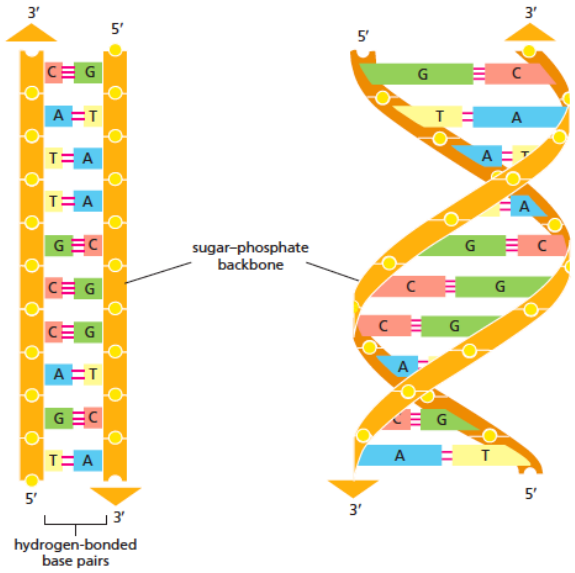


Figure 70 : Composition de la double hélice d'ADN (Alberts *et al.,* 2013, p. 173). Les nucléotides sont liés de façon covalente ensemble en des chaines polynucléotidiques, avec une squelette sucre-phosphate sur lequel les bases (A,C,G, et T) sont fixées. Une molécule ADN est composée de deux chaines polynucléotidiques, appelées « brins d'ADN », et tenues ensemble par des liaisons hydrogènes entre les paires de base. Les flèches sur les brins d'ADN indiquent le sens de chacun des deux brins, qui sont antiparallèles l'un par rapport à l'autre. Bien que l'ADN soit représenté de façon « aplatie » (à gauche) les deux brins d'ADN forment en fait une double hélice.

Allèle : chacune des différentes formes sous lesquelles un gène peut exister et qui contribue à la détermination d'un caractère phénotypique. Pour un même gène, il peut exister de nombreux allèles dans une espèce donnée. Pour chaque gène, un individu hérite d'un allèle paternel et d'un allèle maternel qui peuvent ou non être identiques. S'ils ne sont pas identiques, un allèle peut s'exprimer et l'autre pas, déterminant le caractère. On dit que l'allèle qui s'exprime est dominant et l'autre récessif (voir « dominance » et « récessivité »). Les allèles diffèrent entre eux par mutation mais le terme de mutation n'est pas synonyme de pathologie (voir « mutation »). Une mutation peut n'avoir aucune conséquence sur le phénotype (mutation silencieuse ou polymorphisme). Quand la mutation du gène entraîne une maladie, on parle d'**allèle morbide**.

Anomalie chromosomique : Une altération ou une anomalie ne peut être définie que par rapport à un état défini comme « normal ». La grande stabilité de la formule chromosomique et de la structure globale des chromosomes dans la population humaine ne se vérifie que globalement lorsque les chromosomes sont observés au microscope (caryotype conventionnel). On peut alors déceler des altérations chromosomiques par rapport au caryotype normal.

ARN : Acide ribonucléique : molécule dont la structure est proche de l'ADN, et qui est formée d'un enchaînement précis de quatre types de nucléotides : l'adénine, la guanine, la cytosine et l'uracile. Il existe toute sorte de molécules d'ARN qui peuvent avoir différentes fonctions au sein de la cellule : ARN messager (ARNm), ARN ribosomal (ARNr), ARN de transfert (ARNt). Tous les ARN ne sont pas codants.

ARN messager (ARNm) : un ARN messager est synthétisé à partir d'une séquence d'ADN lors de la transcription qui a lieu dans le noyau. L'ARN messager est ensuite transféré dans le cytoplasme, où il est traduit en une séquence d'acides aminés, selon le principe de correspondance du code génétique.

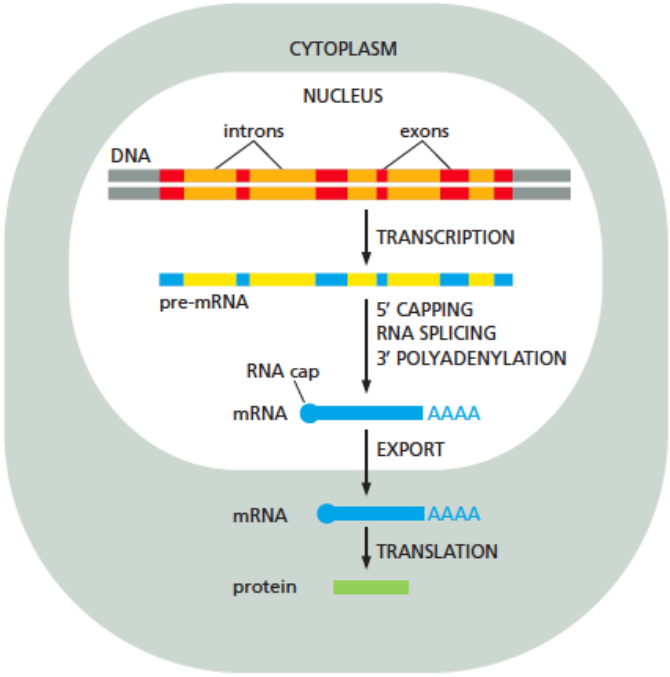


Figure 71 : De l'ADN à l'ARNm et de l'ARNm à la protéine (Alberts *et al.*, 2013, p. 238). Dans les cellules eucaryotes, la transcription de l'ADN produit un ARN pré-messager qui contient à la fois les séquences introns et les séquences exons. Les deux extrémités de cet ARN pré-messager sont modifiées et les introns sont épissés par épissage alternatif. L'ARN messager final est alors transporté du noyau vers le cytoplasme où il est traduit en protéine.

Caractère continu ou quantitatif : caractère dont la variation quantitative dans une population est continue et suit une loi normale, c'est-à-dire une courbe de Gauss et sur lequel l'influence de l'environnement est patente. Exemple : distribution de la taille humaine

Caractère discret ou qualitatif : caractère dont la variation dans la population est discontinue, peut prendre un nombre limité de valeurs et dont l'expression est généralement considérée comme indépendante de l'environnement. Exemple : avoir ou ne pas avoir les yeux bleus.

Chromatine : l'ADN nucléaire présent dans une cellule humaine mesure environ deux mètres, alors que le noyau qui le contient mesure 5 à 7 micromètres. La chromatine est la substance de base des chromosomes eucaryotes, qui permet d'organiser de façon dynamique la forme que prend l'ADN dans le noyau, en compactant ou au contraire en exposant certaines régions de l'ADN, en fonction de l'expression

génétique. La chromatine est composée d'ADN et de protéines (les histones et les non-histones), empaquetés sous la forme de nucléosomes qui représentent le premier niveau de la compaction de l'ADN.

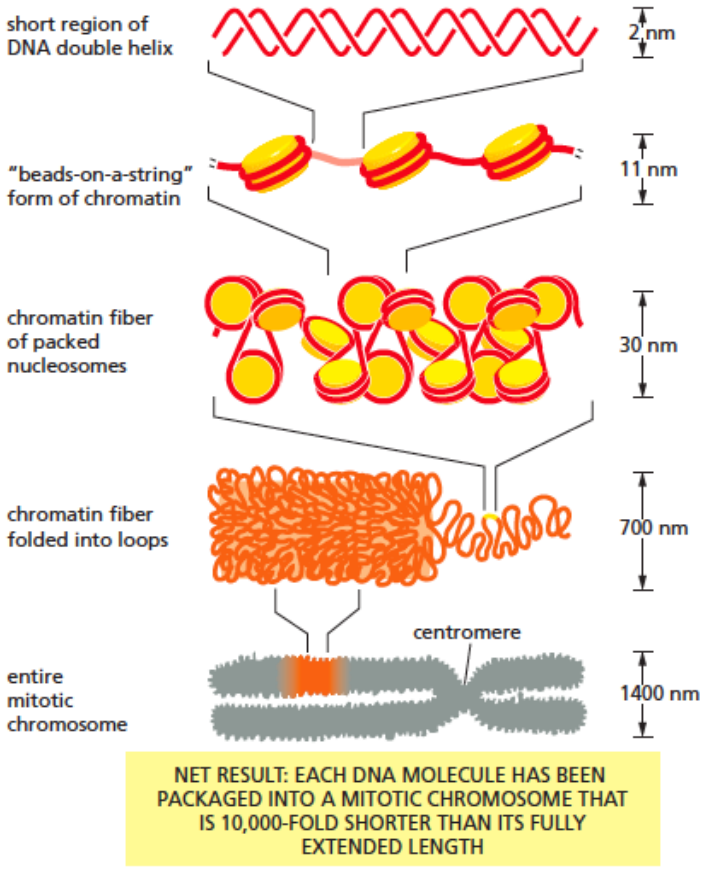

short region of DNA double helix
2 nm
"beads-on-a-string" form of chromatin
11 nm
chromatin fiber of packed nucleosomes
30 nm
chromatin fiber folded into loops
700 nm
centromere
entire mitotic chromosome
1400 nm
NET RESULT: EACH DNA MOLECULE HAS BEEN PACKAGED INTO A MITOTIC CHROMOSOME THAT IS 10,000-FOLD SHORTER THAN ITS FULLY EXTENDED LENGTH

Figure 72 : Les différents niveaux d'empaquetage de l'ADN dans les chromosomes (Alberts *et al.*, 2013, p. 187). L'ADN enroulé autour d'un assemblage de huits molécules d'histones constitue un nucléosome. Les nucléosomes constituent une structure « en collier de perles ». La chromatine est composée d'un ensemble de nucléosomes empaquetés. La fibre de chromatine est repliée en boucles qui constituent le chromosome. Au total, chaque molécule d'ADN a été empaquetée dans un chromosome mitotique qui est 10 000 fois plus petit que la longueur totale de l'ADN.

Chromosome : unité chromatinienne distincte visible au microscope optique lors de la division cellulaire. Chez l'homme, il existe 23 paires de chromosomes (hérités du père et de la mère) dont 22 paires de chromosomes dits autosomiques et une paire de chromosomes sexuels (X et Y). Les femmes possèdent deux chromosomes X et les hommes un chromosome X et un chromosome Y. L'image des chromosomes lors de la division cellulaire constitue le caryotype sur lequel certaines anomalies conduisant à

des maladies génétiques peuvent être détectées, notamment lors d'examens de diagnostic prénatal.

CIM : Classification statistique internationale des maladies et des problèmes connexes de santé. Héritière de la classification des causes de décès de Jacques Bertillon (première édition en 1938), cette classification qui répertoriait uniquement les causes de mortalité a connu un changement majeur en s'intéressant à partir de sa sixième révision en 1946 aux causes de morbidité et donc à la classification des maladies. Ce travail de codage et de classification est alors confié à l'Organisation Mondiale de la Santé (OMS) et elle devient la « Classification statistique internationale des maladies, traumatismes et causes de décès ». L'OMS a brisé le rythme des révisions décennales avec la dernière version de la CIM : étant donné l'importance des modifications à apporter, la CIM-10 n'a vu le jour qu'en 1993 et fait l'objet depuis de constantes mises à jour en ligne. L'arrivée officielle de la CIM-11 est prévue pour 2015.

CNV : *Copy Number Variant* : c'est une des sources de variation génétique dans le génome humain (voir « polymorphisme ») : un segment d'ADN d'au moins un kilobase et dont le nombre de copies est variable en comparaison à un génome de référence. Un CNV peut être simple dans sa structure (comme une duplication simple), ou peut impliquer des gains ou des pertes complexes de séquences homologues en de multiples sites du génome.

Code génétique : ensemble des règles de correspondance (code) découvertes dans les années 1960 et qui permettent la traduction de l'ARN en protéines dans une cellule. A chaque séquence de trois bases consécutives (appelé codon) portées par l'ARN messager, correspond un acide aminé donné. C'est le code génétique qui permet donc la traduction des « messages » codés dans le génome en protéines ayant des fonctions bien précises.

Comorbidité : terme ambigu qui désigne en première intention le fait que deux ou plusieurs maladies sont présentes simultanément chez un même individu. La question décisive est alors de déterminer à quel point cette comorbidité est une coïncidence ou

révèle un lien (facteur d'exposition, mécanisme co-évolutif, mécanisme physiopathologique) commun entre les deux maladies. Par extension, la comorbidité désigne donc la probabilité qu'un malade atteint d'une maladie A soit atteint d'une maladie B à un moment donné de son existence.

Congénitale (maladie) : une maladie congénitale est présente à la naissance qu'elle soit génétique ou non. Ainsi, certaines maladies infectieuses comme la rubéole ou la toxoplasmose par exemple peuvent engendrer des anomalies congénitales. À l'inverse, beaucoup de maladies dites génétiques ne sont pas congénitales et ne s'expriment qu'au cours de la vie.

Délétion chromosomique : une délétion est une perte d'ADN chromosomique dont la taille peut être très variable. Il existe des délétions ponctuelles, d'une seule base, ou bien de grande taille. On parle de microdélétion lorsque cette lacune est à la limite de détection des techniques classiques d'analyse par cytogénétique conventionnelle des chromosomes.

De novo : non hérité des parents. Voir « néomutation ».

Dominant : chacun de nous porte deux allèles (hérités de chacun de ses deux parents) d'un même gène pour un caractère donné. Soit A et B deux allèles possibles pour un même gène donné, l'allèle A est dit dominant sur l'allèle B si les phénotypes associés au génotype homozygote AA et hétérozygote AB sont identiques. L'allèle B est dit alors récessif. Si le phénotype d'un sujet AB est intermédiaire entre ceux résultant de AA et de BB, les allèles A et B sont dits semi-dominants. Si le sujet AB exprime à la fois ce qui est observé pour le génotype AA et pour le génotype BB, les allèles sont dits codominants (c'est le cas des groupes sanguins A et B).

Drépanocytose (ou anémie falciforme) : maladie autosomique récessive, liée à des mutations sur le gène HbS codant pour la production de l'hémoglobine et qui est caractérisée par des globules rouges en forme de faucille. Les manifestations cliniques de la drépanocytose sont de trois grands types : anémies hémolytiques (destruction

des globules rouges anormaux), susceptibilité accrue aux infections bactériennes et accidents ischémiques vaso-occlusifs dus aux conflits entre les petits vaisseaux et les globules rouges anormaux qui sont peu déformables.

Epigénétique : changements fonctionnels du génome qui peuvent être hérités, mais n'entraînent pas de changements dans la séquence de l'ADN.

Épissage : procédé par lequel l'ARN messager se transforme et se raccourcit, excluant les séquences correspondant aux introns et fusionnant les exons. L'épissage s'effectue grâce à des séquences spécifiques situées aux extrémités 5'et 3' de chaque exon. L'épissage est une des étapes de la maturation de l'ARN messager (après la transcription) chez les eucaryotes. Elle intervient avant le passage de l'ARNm dans le compartiment cytoplasmique.

Épissage alternatif : Mécanisme par lequel des ARN messagers matures composés d'exons différents sont produits à partir du même ARN messager. Ainsi, pour un même gène, il y aura plusieurs ARNm différents qui donneront naissance à des protéines différentes. Les différents ARNm peuvent être produits dans des cellules différentes ou à des temps différents de la vie cellulaire. Ce phénomène d'économie pour la cellule est omniprésent chez tous les eucaryotes pluricellulaires.

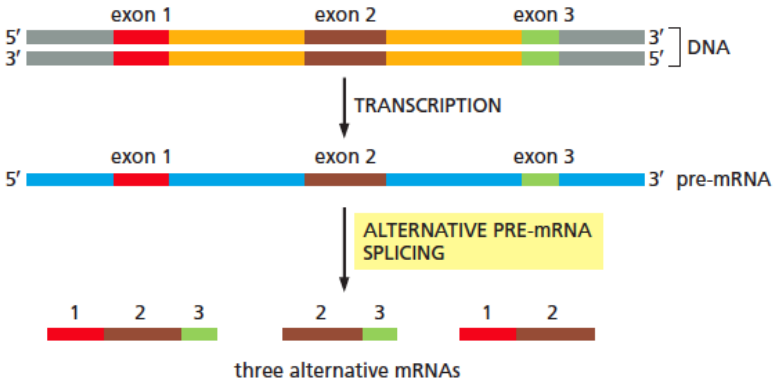


Figure 73 : Épissage alternatif (Alberts *et al.*, 2013, p. 236). Certains ARN pré-messagers subissent un épissage alternatif. Alors que tous les exons sont présents dans un ARN pré-messager, certains exons peuvent être exclus de l'ARN messager final. Dans cet exemple, trois des quatre ARNm possibles sont produits (manque la séquence d'exons 1-3). Les extrémités 3' et 5' des ARN messagers ne sont pas représentées.

Epistasie : désigne l'interaction existant entre deux ou plusieurs gènes (interactions GxG). Il y a épistasie lorsqu'un ou plusieurs gènes empêchent l'expression d'un ou plusieurs gènes situés sur d'autres loci. En génétique mendélienne, qui s'intéresse aux caractères discrets ou qualitatifs, une relation d'épistasie désigne en général la suppression d'un phénotype par un autre gène de la même voie métabolique. En génétique quantitative, qui s'intéresse aux caractères continus ou quantitatifs, l'épistasie désigne la part de variance génétique qui ne peut être expliquée ni par les effets additifs des allèles en présence, ni par les effets de dominance.

Exome : l'exome est la partie du génome dont la séquence est traduite en protéines, celle qui est le plus directement liée au phénotype, aux maladies génétiques, en terme médical. L'exome humain représente environ 1,5% de son ADN. Diverses sociétés commerciales proposent de séquencer et d'analyser l'exome d'individus pour rechercher les variations qui seraient à l'origine des maladies génétiques.

Expressivité variable : le même allèle morbide peut causer des symptômes cliniques variables d'un individu à un autre pour une même maladie donnée.

Gène : le terme de gène a été inventé en 1909 par le botaniste danois Wilhelm Johannsen pour décrire ce que les parents transmettent à leur descendance et qui exprime un trait particulier de leur phénotype. Le gène de la génétique dite classique a donc été longtemps un support hypothétique de l'hérédité dont on ne connaissait pas la nature matérielle. Oswald Avery, Colin MacLeod et Maclyn McCarty ont démontré en 1944 que le support primaire du gène est l'ADN. Chez les mammifères, l'homme en particulier, on hérite de deux copies de chaque gène (une copie paternelle et une copie maternelle) qui peuvent ne pas être strictement identiques. On parle alors de deux formes alléliques ou de deux allèles du gène considéré. La définition du gène est aujourd'hui sujette à de nombreuses controverses, Griffiths et Stotz ont ainsi démontré la coexistence de trois concepts différents de gène au sein de la littérature biomédicale contemporaine : le « gène instrumental », le « gène nominal » et le « gène post-génomique ». Le gène instrumental est avant tout un instrument formel qui constitue une unité de transmission des caractères mendéliens et qui est donc

fortement utilisé par les biologistes évolutionnistes et par la génétique des populations. Le gène nominal est véritablement un héritage de la biologie moléculaire : il est défini par la linéarité de sa séquence, c'est-à-dire par les outils du génie génétique qui ont permis de l'identifier et de l'annoter. Le gène post-génomique est l'héritier des découvertes des années 1970 et du Projet Génome Humain.

Génétique : science de l'hérédité, la génétique étudie la transmission de traits phénotypiques chez les individus d'une même espèce, ainsi que la relation entre génotype et phénotype. L'étude de l'hérédité de caractères simples (chez les plantes) par le moine tchèque Grégor Mendel dans les années 1860 lui a permis d'établir des lois régissant cette transmission (génétique formelle ou mendélienne). Mais l'invention en 1905 du terme « génétique » revient au biologiste William Bateson (1861-1926).

Génétique inverse : La génétique dite « classique » s'appuie sur une approche aussi appelée « physiologique » : partant de ce que l'on connaît de la physiologie de la maladie (symptômes, fonction des protéines impliquées, etc.), on cherche à retrouver le gène ou les gènes porteur(s) des mutations impliquées dans la maladie. Au contraire, la génétique inverse ou clonage positionnel cherche à identifier le gène en identifiant sa position dans le génome (voir Annexe 1).

Génome : ensemble de l'information génétique d'un individu, structure et organisation moléculaire qui spécifie cette information. Le génome humain diploïde est composé de 6 milliards de paires de bases réparties sur 23 paires de chromosomes. Seul 1% du génome humain, soit 60 millions de paires de bases, constitue les parties codantes des 23 000 paires de gènes qui seront traduits en protéines. L'ensemble des parties codantes des 23 000 gènes constitue l'exome. Des efforts intenses de recherche sont réalisés pour identifier les rôles des 99% restant du génome humain ; c'est en particulier le projet du consortium international ENCODE (Encyclopedia of DNA éléments). Une part de ces régions non codantes a un rôle majeur dans la régulation de l'expression des gènes.

Génotype : par rapport à un ou plusieurs caractères héréditaires, le génotype est l'ensemble des allèles que porte un individu pour les gènes susceptibles d'agir sur ce ou ces caractères. (voir phénotype).

Hérédité mendélienne : les maladies mendéliennes sont des maladies dont la transmission suit les lois de Mendel, qui postulent qu'un caractère est déterminé par deux formes différentes (correspondant aux allèles) d'un locus donné, chacune de ces formes étant transmise par l'un des parents. Depuis la découverte de la nature matérielle du gène, la maladie mendélienne et la maladie monogénique sont devenues synonymes puisqu'on considère qu'une maladie monogénique est déterminée par un seul gène. On distingue en général l'hérédité mendélienne liée aux chromosomes autosomiques (maladie autosomique dominante et maladie autosomique récessive) et l'hérédité mendélienne liée aux chromosomes X (maladie dominante liée à l'X, maladie récessive liée à l'X). Le cas des maladies liées à l'Y est un peu particulier, puisque d'une part ces maladies sont extrêmement rares et que d'autre part, par définition il ne peut pas y avoir de phénomène de dominance ou de récessivité dans les maladies liées à l'Y. Par analogie avec l'hérédité mendélienne liée à l'X, on considère cependant que l'hérédité liée à l'Y fait partie de l'hérédité mendélienne.

Hérédité non conventionnelle : toutes les formes d'hérédité non mendélienne comprenant entre autres l'hérédité oligogénique, l'hérédité polygénique, l'hérédité mitochondriale, les phénomènes d'empreinte parentale, etc.

Héritabilité au sens strict et au sens large : si le terme a été forgé par Jay Lush en 1940, le concept d'héritabilité est largement tributaire des travaux de Fisher sur la variation des caractères continus et de son hypothèse selon laquelle on peut décomposer de façon additive la variance phénotypique (V_P) d'un caractère en sa variance génotypique (V_G) et sa variance environnementale (V_E) : $V_P = V_G + V_E$. L'héritabilité est alors définie comme la part de variance phénotypique (V_P) due à la variance génotypique héritée (V_G).

Chez les êtres humains, lors de la reproduction sexuée, la contribution de chaque parent se limite à la transmission d'un gamète haploïde (contenant la moitié

des chromosomes parentaux), issu de la méiose et résultant d'une recombinaison. Les interactions génétiques sont donc rompues et chaque allèle est transmis indépendamment. Ainsi, dans un gamète haploïde, la variation due à l'effet de dominance et la variation due à l'effet des interactions épistatiques (interactions gène x gène) ne sont pas transmises à la descendance – seule la variance additive est transmise.

L'héritabilité au sens large décompose la variance génotypique (V_G) en prenant en compte la variance additive (V_A), la variance due à l'effet de dominance (V_D) et la variance due aux interactions épistatiques (V_I). L'héritabilité au sens large (notée H^2) mesure donc la part de variance phénotypique dans une population due aux différences génétiques, indépendamment de l'origine de cette différence génétique. On peut donc écrire : $H^2 = (V_A + V_D + V_I) / V_P$. En revanche, l'héritabilité au sens strict (notée h^2) ne prend en compte que la variance additive, c'est-à-dire uniquement l'effet additif de plusieurs gènes impliqués dans la détermination des caractères. On peut donc écrire : $h^2 = V_A / V_P$.

Hétérogénéité allélique (phénomène « plusieurs allèle, une même maladie ») : le fait que plusieurs mutations d'un même gène (c'est-à-dire plusieurs allèles morbides) peuvent être associées à une même maladie. Par exemple, on a répertorié plus de 500 mutations du gène PAH pouvant causer la phénylcétonurie, et plus de 1000 mutations du gène CFTR sont associées à la mucoviscidose.

Hétérogénéité génétique (phénomène « plusieurs gènes, une même maladie ») : plusieurs gènes peuvent causer une même maladie. Par exemple, plus de quarante gènes seraient impliqués dans les rétinites pigmentaires.

Hétérozygote : individu qui porte deux allèles différents (chacun de ces allèles ayant été hérité de l'un des parents) d'un même gène à un locus donné pour un caractère donné.

Homozygote : individu qui porte deux allèles identiques (hérités de ses deux parents) à un locus donné pour un caractère donné.

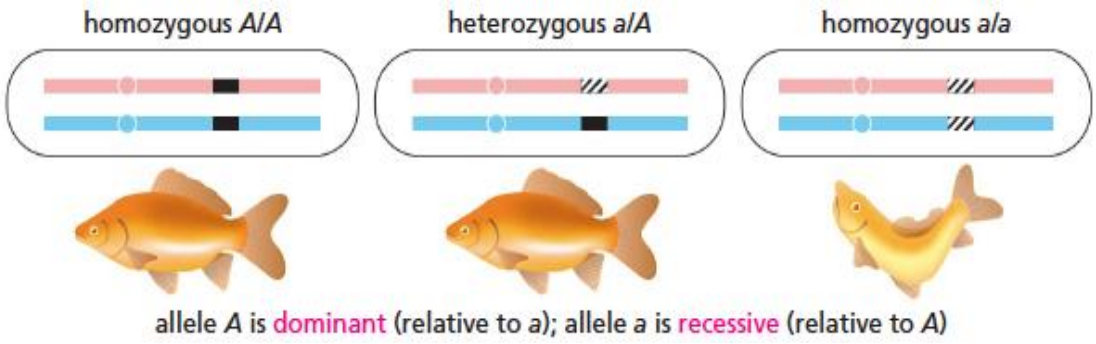


Figure 74 : Hétérozygotie et homozygotie, dominance et récessivité (Alberts *et al.,* 2013, p. 666). Dans cet exemple, l'allèle *A* est dominant (relativement à l'allèle *a*) puisque l'homozygote *A/A* et l'hétérozygote *a/A* ont le même phénotype. L'allèle *a* est dit récessif, relativement à *A*. Dans cet exemple, le phénotype de l'hétérozygote est le même que celui de l'un des homozygotes. Lorsque le phénotype hétérozygote est différent des deux phénotypes homozygotes, on dit que les deux allèles sont co-dominants.

Hypothèse variant commun, maladie commune : hypothèse développée à la fin des années 1990 (Lander, 1996 ; Reich et Lander, 2001) et qui a dominé la génétique des maladies complexes. Selon cette hypothèse, les maladies communes ne sont pas dues à une infinité de variants alléliques à faible effet sur le phénotype (ce qui correspond au modèle infinitésimal de Fischer) mais à un nombre modeste de variants communs (dont la fréquence dans la population serait supérieure à 1%). Aucun de ces variants ne suffit à provoquer la maladie (ils sont rarement délétères) ce qui expliqueraient leur maintien dans la population générale (s'ils étaient délétères, on pourrait s'attendre à ce que la sélection naturelle aboutisse à leur élimination). Paradoxalement, même les principaux défenseurs de cette hypothèse ont toujours reconnu que cette hypothèse manquait de preuve empirique (Reich et Lander, 2001) – mais c'est l'hypothèse la plus logique pour expliquer à la fois le caractère commun des maladies, la complexité de la détermination génétique de ces maladies et le maintien de ces variants dans la population générale. Si les variants responsables des maladies communes étaient au contraire rares mais avec un effet très important sur le phénotype global, les maladies communes seraient moins fréquentes et la présence d'un allèle muté dans une famille donnerait un modèle de transmission très proche de l'hérédité mendélienne. L'hypothèse « variant commun-maladie commune » a par ailleurs bénéficié de

quelques exemples bien établis, comme le variant ApoE dans la maladie d'Alzheimer, qui ont contribué à en faire un modèle admis par tous.

Interactome humain : l'interactome humain désigne l'ensemble des interactions moléculaires (interactions gène-gène, interactions gène-facteur de transcription, interactions protéines-protéines, etc.) ayant lieu dans un être humain.

Locus : site physique où se situe un gène donné sur un chromosome donné.

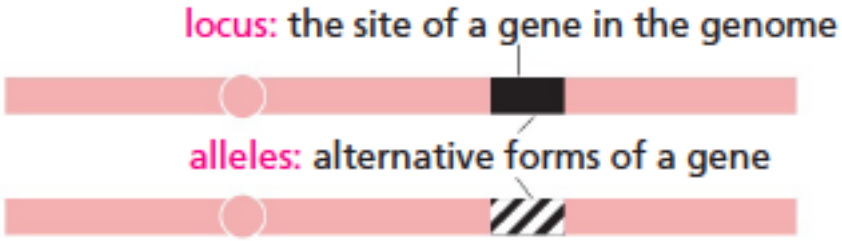


Figure 75 : Locus et allèles (Alberts *et al.,* 2013, p. 666): le locus d'un gène est le site d'un gène dans le génome (représenté par le trait vertical). Les allèles sont les formes alternatives d'un gène : dans cet exemple, il y a deux allèles, deux formes du même gène, représentés par le carré noir et par le carré hachuré.

Maladie monogénique : une maladie est dite monogénique si l'apparition du phénotype est déterminée par un seul gène.

Maladie oligogénique : une maladie est dite oligogénique si l'apparition du phénotype est déterminée par l'action conjointe de quelques gènes et de facteurs environnementaux.

Maladie polygénique : une maladie est dite polygénique si l'apparition du phénotype est déterminée par l'action conjointe d'un grand nombre de gènes (chacun de ces gènes exerçant un effet mineur sur le phénotype final) et de facteurs environnementaux.

Microbiome humain : c'est la communauté écologique des micro-organismes commensaux, symbiotiques ou pathogènes qui partagent le corps humain. On

distingue parfois « microbiome » et « microbiotes », le premier terme désignant l'ensemble des génomes des micro-organismes qui résident dans une niche environnementale donnée, le second désignant alors les micro-organismes eux-mêmes.

Mutation : variation de la séquence d'ADN à un endroit particulier du génome d'un individu.

- Type de variation : Cette variation peut être ponctuelle, c'est-à-dire n'affecter qu'une des trois milliards de bases de l'ADN, ou impliquer des régions plus ou moins grandes du génome (substitution ; délétion ; addition ; remaniement chromosomique ; insertion).
- Origine de la mutation : Une mutation peut être héréditaire : un des parents ou les deux parents transmettent la mutation à leur descendance. Une mutation peut aussi survenir au moment de la formation d'un gamète parental (mutation germinale ou de novo ou néomutation), auquel cas l'embryon sera porteur d'une mutation qu'aucun des deux parents n'avait dans son patrimoine génétique. Enfin, une mutation peut avoir lieu au cours de la vie dans les cellules non germinales (mutation somatique), elles ne seront jamais transmises à la descendance.
- Locus de la mutation : une mutation peut avoir lieu dans les cellules germinales ou dans les cellules somatiques, dans l'ADN nucléaire ou dans l'ADN mitochondrial, elle peut concerner une région codante ou non codante, etc.
- Effet d'une mutation : On parle ainsi de « mutation silencieuse » lorsque la mutation d'un allèle n'a pas d'effet sur l'expression du phénotype et on appelle cet allèle un « polymorphisme ». Une mutation peut aussi entrainer un gain de fonction, auquel cas on parle d' « allèle hypermorphe », une nouvelle fonction (« allèle néomorphe »), une perte de fonction totale (« allèle amorphe ») ou partielle (« allèle hypomorphe »), inhiber le produit de l'allèle sauvage (« allèle dominant négatif »). Une mutation peut être conditionnelle, c'est-à-dire que son effet n'est perçu qu'en fonction d'une condition environnementale particulière (par exemple, les mutations thermosensibles). Une mutation peut

avoir plusieurs effets, causant différents phénotypes (« mutation à effet pléiotrope »).

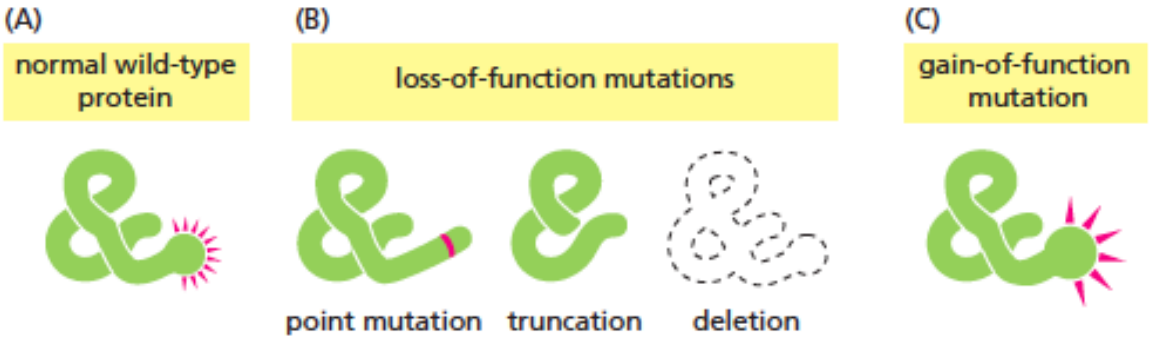


Figure 76 : Différents types de mutations ont différentes conséquences sur la fonction d'une protéine (Alberts *et al.,* 2013, p. 666). (A) Dans cet exemple, la protéine normale (« protéine sauvage ») a une fonction spécifique dénotée par les rayons rouges. (B) Différentes mutations de type « perte de fonction » diminuent ou inhibent complètement l'activité de la protéine. (C) Une mutation gain de fonction accroît cette activité (ce qui est dénoté par l'épaisseur des rayons rouges) ou augmente le nombre de protéines produites (non représenté ici).

Néomutation : une mutation dans un gène donné peut survenir de façon accidentelle dans les gamètes de l'un des deux parents ou très tôt après la constitution du zygote. Elle n'est donc pas héritée des parents.

OMIM : *Online Mendelian Inheritance on Man* : base de données de référence en ligne de la génétique médicale, qui répertorie toutes les associations connues entre un variant allélique et une maladie et qui est administrée depuis 1987 par la Bibliothèque Nationale de médecine américaine. Elle a été créée en 1966, par Victor McKusick, sous forme d'un catalogue papier alors intitulé « Hérédité mendélienne chez l'homme. Catalogue des phénotypes autosomiques dominants, autosomiques récessifs et X-liés » et ne concernait d'abord que les maladies monogéniques. Mais depuis 1996, elle s'est ouverte aux maladies polygéniques.

Pénétrance : fréquence avec laquelle le porteur d'un allèle exprime le caractère associé à cet allèle. Lorsque la pénétrance est incomplète, cela signifie que des individus porteurs d'un même génotype n'exprimeront pas le même phénotype. La

pénétrance repose sur une combinaison de facteurs à la fois génétiques et environnementaux, qui, dans la plupart des cas, reste à identifier.

Phénylcétonurie : longtemps considérée comme une maladie-star de la génétique médicale, la phénylcétonurie entraine une accumulation de phénylalanine dans le sang qui peut être la cause d'un retard mental. Cette hyperphénylalanisme est généralement due à une mutation autosomique récessive du gène PAH qui code pour la phénylalanine hydroxylase, une enzyme qui transforme la phénylalanine en tyrosine. Dans d'autres cas plus rares, la maladie est causée par des mutations dans le gène BH4 codant pour le co-facteur de la phénylalanine hydroxylase. L'apparition de la maladie peut être empêchée, à condition de suivre un régime sans phénylalanine. Cette maladie fait l'objet d'un dépistage à la naissance systématique en France.

Phénotype : Caractères observables d'un individu considéré soit sous un aspect particulier, soit sous la totalité de ses aspects. Le phénotype dépend à la fois du génotype et du milieu". Le phénotype relève de traits morphologiques ou physiologiques, voire comportementaux. La relation entre phénotype et génotype n'est pas univoque puisque, par exemple, il n'y a pas d'identité phénotypique totale entre deux jumeaux vrais (jumeaux monozygotes).

Pléiotropie (phénomène « un gène, plusieurs phénotypes ») : ce phénomène, lié entre autres à l'épissage alternatif, désigne le fait qu'un gène puisse être associé à de multiples phénotypes. Par exemple, le gène lamine A/C est associée à plus de onze maladies différentes, dont la progéria et la maladie de Charcot-Marie-Tooth.

Polymorphisme génétique : Variation entre individus dans la séquence de gènes. Dans une population on parle alors de polymorphisme génique lorsqu'au moins deux allèles d'un gène sont présents et à une fréquence égale ou supérieure à un pour cent. Ces variations qui rendent compte des différents allèles dans une population ne sont pas pathogènes la plupart du temps. Il existe différents types de polymorphismes génétiques, selon que la variation touche une région codante ou non codante du

génome, concerne un seul nucléotide (voir « SNP ») ou le nombre de copies d'un même variant allélique (voir « CNV »).

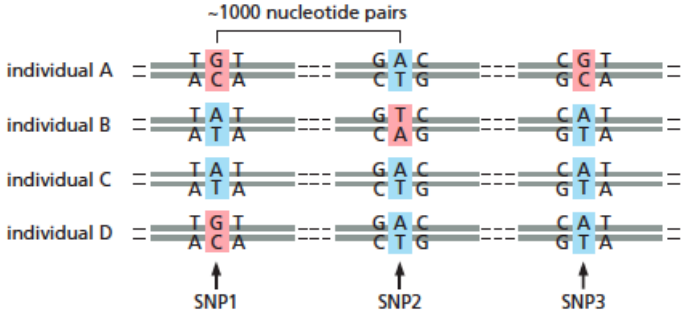


Figure 77 : Un exemple de polymorphisme génétique : les polymorphismes à un seul nucléotide (SNP) (Alberts *et al.,* 2013, p. 672). Les SNP sont des sites du génome où il est fréquent que des individus appartenant à une population donnée soient porteurs de nucléotides différents. La plupart de ces variations dans le génome humain n'affectent pas significativement la fonction du gène.

Porteur sain : le terme « porteur sain » a deux sens différents quoique proches selon qu'il est utilisé dans le contexte des maladies infectieuses ou des maladies génétiques. Dans le cas des maladies infectieuses, ce terme désigne un individu qui est infecté par un micro-organisme pathogène mais qui ne développe aucun symptôme de la maladie. Le cas le plus connu est celui de Mary Mallon, surnommée « Mary Thyphoïde », pour avoir été le premier exemple de porteur sain de la fièvre typhoïde identifié aux États-Unis. En génétique médicale, le terme de porteur sain désigne d'abord un individu hétérozygote pour une maladie génétique récessive : cet individu n'a donc généralement aucun symptôme de la maladie (en dehors de circonstances particulières : certains facteurs environnementaux peuvent déclencher des symptômes chez un patient hétérozygote atteint de drépanocytose). Plus généralement, le terme de porteur sain peut être utilisé pour désigner des individus chez qui la pénétrance d'une maladie donnée est incomplète : ils ne développent pas de symptômes d'une maladie, alors qu'ils ont un génotype associé dans la population générale à l'apparition de la maladie.

Projet Génome Humain : projet scientifique international entrepris dans les années 1990 et dont l'objectif était de parvenir à la séquence complète de l'ADN du génome humain (avril 2003).

Problème de l'héritabilité perdue : L'héritabilité permet de mesurer la part de variance génétique dans la variance phénotypique d'un trait dans une population donnée. De façon indépendante, plusieurs méthodes ont été successivement mises au point pour identifier quels sont les variants alléliques responsables de la variance génétique dans la variation phénotypique d'une maladie donnée. Or lorsque l'on confronte l'estimation de l'héritabilité d'une maladie d'une part et la variance additive des variants alléliques impliqués dans la maladie en question, il y a une discordance significative. Autrement dit, une fraction de l'héritabilité estimée d'un grand nombre de traits ou de maladies reste inexpliquée par la variance additive des variants alléliques impliqués dans le trait ou la maladie en question : c'est le problème de l'héritabilité perdue.

Récessif : chacun de nous porte deux allèles (hérités de chacun de ses deux parents) d'un même gène pour un caractère donné. Soit A et B deux allèles possibles pour un même gène donné, l'allèle A est dit dominant sur l'allèle B si les phénotypes associés au génotype homozygote AA et hétérozygote AB sont identiques. L'allèle B est dit alors récessif. Si le phénotype d'un sujet AB est intermédiaire entre ceux résultant de AA et de BB, les allèles A et B sont dits semi-dominants. Si le sujet AB exprime à la fois ce qui est observé pour le génotype AA et pour le génotype BB, les allèles sont dits codominants (c'est le cas des groupes sanguins A et B).

Séquence, séquençage de l'ADN : on sait aujourd'hui lire la succession (séquence) des quatre bases moléculaires qui se suivent le long des molécules d'ADN présentes dans le génome. On a pu séquencer ou établir la séquence de l'AND présent dans le génome humain pour la première fois au début de l'année 2000. Cette première séquence a représenté 13 ans de travail pour un montant de l'ordre de 3 milliards de dollars. Aujourd'hui, la même chose peut être réalisée en quelques jours pour moins de 1000 dollars.

SNP : Single Nucleotide Polymorphism : type de polymorphisme (voir « polymorphisme ») de l'ADN dans lequel deux chromosomes diffèrent sur un segment donné par une seule paire de bases. Dans deux génomes humains tirés au hasard, 99,9% de la séquence d'ADN est identique. Les 0,1% restants contiennent des variations de séquence dont le type le plus commun est le polymorphisme pour un nucléotide (SNP). Un autre type de variation est le nombre de copies d'un variant allélique (voir « CNV »). Les SNP sont stables, très abondants et distribués uniformément dans tout le génome, c'est pourquoi ils sont utilisés comme marqueurs génétiques (voir « marqueurs génétiques »).

Figure 78 : Comparaison des SNP chez des individus sains et des individus malades (Alberts *et al.*, 2013, p. 675). Les motifs des SNP de deux ensembles d'individus sont comparés : un ensemble d'individus sains qui servent de contrôle et un ensemble d'individus affectés par une maladie commune. Un segment de chromosome typique est montré. Pour la plupart des sites polymorphiques de ce segment, le fait qu'un individu ait un type de SNP (barre verticale rouge) ou un autre (barre verticale bleue) n'a pas de conséquences. Cependant, dans la portion gris foncé du chromosome, on remarque que la plupart des individus normaux ont des variants SNP bleus alors que la plupart des individus malades ont des variants SNP rouges. Ceci suggère que cette région contient ou est proche d'un gène qui est lié génétiquement à ces variants SNP rouges et qui prédisposent à la maladie.

Trisomie 21 ou syndrome de Down : exemple particulier d'aneuploïdie dans lequel un chromosome est présent, dans son intégralité, en trois copies au lieu de deux dans chaque cellule de l'organisme. Deux mécanismes principaux peuvent expliquer la présence de ce chromosome surnuméraire. Dans 95% des cas, la trisomie est dite

« libre » : les chromosomes homologues ne se sont pas séparés pendant la méiose de la gamétogénèse. Dans les autres cas, la trisomie est dite « non libre » : il s'agit de la fusion de deux chromosomes 21 par translocation. La trisomie de certains chromosomes n'est pas compatible avec la vie. Dans le cas de la trisomie du chromosome 21, dite trisomie 21, ou syndrome de Dawn, la vie fœtale peut, dans une grande proportion de cas, se poursuivre jusqu'à la naissance et au-delà. Les anomalies du développement et les symptômes associés, notamment en termes de retard d'apprentissage et de handicap intellectuel, s'expriment de manière variable selon les cas d'une manière qui, à ce jour, ne peut pas être prédite.

Variance : mesure statistique servant à caractériser la dispersion des valeurs d'une population autour de la moyenne des valeurs de cette population.

Variance phénotypique : la variance phénotypique désigne la dispersion des variations possibles d'un même phénotype dans la population. C'est un paramètre accessible à l'observation : il suffit de faire le relevé quantitatif d'un caractère dans une population donnée.

Variant allélique commun : variant allélique dont la fréquence dans la population est supérieure à 1%.

Variant allélique rare : variant allélique dont la fréquence dans la population est inférieure à 1%.

Variabilité interindividuelle : ensemble des variations de l'expression d'une maladie, que ce soit du point de vue des causes de la maladie, des symptômes, de l'évolution ou des traitements, entre les individus d'une population donnée.

Variabilité intra-individuelle : ensemble des variations de l'expression d'une maladie, que ce soit du point de vue des causes de la maladie, des symptômes, de l'évolution ou des traitements, chez un même individu au cours de sa vie.

Bibliographie

ABEL L. et CASANOVA J. L., 2000, « Genetic predisposition to clinical tuberculosis: bridging the gap between simple and complex inheritance », *American Journal of Human Genetics*, vol. 67, no. 2, p. 274.

ABEL L. *et al.,* 1989, « Genetic susceptibility to leprosy on a Caribbean Island: linkage analysis with five markers »*, International journal of leprosy and other mycobacterial diseases: official organ of the International Leprosy Association*, vol. 57, no. 2, p. 465–471.

ABEL L. et DEMENAIS F., 1988, « Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. », *American Journal of Human Genetics*, vol. 42, no. 2, p. 256–266.

ABEL L. et DESSEIN A.J., 1997, « The impact of host genetics on susceptibility to human infectious diseases », *Current Opinion in Immunology*, vol. 9, no. 4, p. 509–516.

ABEL L. *et al.,* 1995, « Complex segregation analysis of leprosy in Southern Vietnam »*, Genetic Epidemiology*, vol. 12, no. 1, p. 63–82.

ALBERTS B. *et al., 2013, Essential cell biology*, Fourth edition., New York, NY, Garland Science.

ALCAÏS A., ABEL L. et CASANOVA J. L., 2009, « Human genetics of infectious diseases: between proof of principle and paradigm », *The Journal of Clinical Investigation*, vol. 119, no. 9, p. 2506–2514.

—, 2010, « Human genetics of infectious diseases », *Vogel and Motulsky's Human Genetics - Problems and Approaches*, M. Speicher, S. E. Antonarakis and A. G. Motulsky éds., Fourth Edition., Heidelberg ; New York, Springer, p. 403–416.

ALCAÏS A., FIESCHI C., *et al.,* 2005, « Tuberculosis in children and adults: two distinct genetic diseases »*, The Journal of Experimental Medicine*, vol. 202, no. 12, p. 1617–1621.

ALCAÏS A., MIRA M., *et al.,* 2005, « Genetic dissection of immunity in leprosy »*, Current Opinion in Immunology*, vol. 17, no. 1, p. 44–48.

ALCAÏS A., QUINTANA-MURCI L., *et al.,* 2010, « Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? »*, Annals of the New York Academy of Sciences*, vol. 1214, no. 1, p. 18–33.

ALTARE F. *et al.,* 1998, « Mendelian susceptibility to mycobacterial infection in man »*, Current Opinion in Immunology*, vol. 10, no. 4, p. 413–417.

ALTER A. *et al.,* 2011, « Leprosy as a genetic disease »*, Mammalian Genome*, vol. 22, no. 1, p. 19–31.

ANDERSON W., 2004, « Natural histories of infectious disease: ecological vision in twentieth-century biomedical science », *Osiris*, vol. 19, , p. 39–61.

ANTMAN E., WEISS S. et LOSCALZO J., 2012, « Systems pharmacology, pharmacogenetics, and clinical trial design in network medicine », *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, vol. 4, no. 4, p. 367–383.

ARRELL D. K. et TERZIC A., 2012, « Systems Proteomics for Translational Network Medicine », *Circulation: Cardiovascular Genetics*, vol. 5, no. 4, p. 478–478.

ASCHARD H. *et al.,* 2012, « Challenges and opportunities in genome-wide environmental interaction (GWEI) studies »*, Human Genetics*, vol. 131, no. 10, p. 1591–1613.

AUFFRAY C., CHEN Z. et HOOD L., 2009, « Systems medicine: the future of medical genomics and healthcare », *Genome Medicine*, vol. 1, no. 1, p. 2.

DE BACKER P., DE WAELE D. et VAN SPEYBROECK L., 2010, « Ins and outs of systems biology vis-à-vis molecular biology: continuation or clear cut? », *Acta biotheoretica*, vol. 58, no. 1, p. 15–49.

BADANO J. L. et KATSANIS N., 2002, « Beyond Mendel: an evolving view of human genetic disease transmission », *Nature Reviews Genetics*, vol. 3, no. 10, p. 779–789.

BADGER J. H., NG P. C. et VENTER J. C., 2011, « The human genome, microbiomes, and disease », *Metagenomics of the Human Body*, p. 1–14.

BAKER A., 2013, « Complexity, networks and non-uniqueness », *Foundations of Science*, vol. 18, , p. 687–705.

BARABÁSI A. L., 1999, « Emergence of Scaling in Random Networks », *Science*, vol. 286, no. 5439, p. 509–512.

—, 2007, « Network medicine—from obesity to the “diseasome” », *New England Journal of Medicine*, vol. 357, no. 4, p. 404–407.

—, 2012, « Network science: Luck or reason », *Nature*, vol. 489, no. 7417, p. 507–508.

BARABÁSI A. L. et BONABEAU E., 2003, « Scale-Free Networks », *Scientific American*, vol. 288, , p. 50–59.

BARABÁSI A. L., GULBAHCE N. et LOSCALZO J., 2011, « Network medicine: a network-based approach to human disease », *Nature Reviews Genetics*, vol. 12, no. 1, p. 56–68.

BARABÁSI A. L. et OLTVAI Z. N., 2004, « Network biology: understanding the cell’s functional organization », *Nature Reviews Genetics*, vol. 5, , p. 101–113.

BASSAGLIA Y., 2013, *Biologie cellulaire*, Paris, Maloine.

BECHTEL W. et ABRAHAMSEN A., 2005, « Explanation: a mechanist alternative », *Studies in history and philosophy of biological and biomedical sciences*, vol. 36, no. 2, p. 421–441.

BECHTEL W. et RICHARDSON R. C., 1993, *Discovering complexity: decomposition and localization as strategies in scientific research*, Princeton, N.J, Princeton University Press.

BECKER K. G. *et al.,* 2004, « The Genetic Association Database »*, Nature Genetics*, vol. 36, no. 5, p. 431–432.

BECKMANN J. S., ESTIVILL X. et ANTONARAKIS S. E., 2007, « Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability », *Nature Reviews Genetics*, vol. 8, no. 8, p. 639–646.

BERENSON R. A., 2004, « Lumpers and splitters: different approaches to understanding vatiations research », *Health Affairs*, Supplement Variation, p. VAR98–VAR103.

BIRD A., 2007, « Perceptions of epigenetics », *Nature*, vol. 447, no. 7143, p. 396–398.

BLAU N. *et al.,* 2011, « Diagnosis, classification, and genetics of phenylketonuria and tetrahydrobiopterin (BH4) deficiencies »*, Molecular Genetics and Metabolism*, vol. 104, , p. S2–S9.

BOEHNKE M., 1994, « Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. », *American Journal of Human Genetics*, vol. 55, no. 2, p. 379–390.

BOORSE C., 1975, « On the Distinction Between Disease and Illness », *Philosophy and Public Affairs*, vol. 5, no. 1, p. 49–68.

—, 1976, « What a Theory of Mental Health should be », *Journal for the Theory of Social Behaviour*, vol. 6, no. 1, p. 61–84.

—, 1977, « Health as a Theoretical Concept », *Philosophy of Science*, vol. 44, no. 4, p. 542–573.

—, 1987, « Concepts of health », *Health Care Ethics: An Introduction*, D. Van De Veer et T. Regan éds., Philadelphia, Temple University Press, p. 359–393.

—, 1997, « A Rebuttal on Health », *What Is Disease?*, J. M. Humber et R. F. Almeder éds., Humana Press, p. 1–134.

—, 2011, « Concepts of Health and Disease », *Philosophy of Medicine*, First edition., Amsterdam ; Boston, Elsevier (Hanbook of philosophy of science), v.16.

BOTSTEIN D. et RISCH N., 2003, « Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease », *Nature Genetics*, vol. 33, no. 3s, p. 228–237.

BOWKER G. C. et STAR S. L., 2000, *Sorting things out : classification and its consequences*, Cambridge, Mass., MIT Press.

BRAILLARD P.A., 2009, *Enjeux philosophiques de la biologie des systèmes*, Université Panthéon-Sorbonne, Paris, France.

BRIDGES C. B., 1936, « The Bar "gene": a duplication », *Science*, vol. 83, no. 2148, p. 210–211.

BRIGANDT I., 2013, « Systems biology and the integration of mechanistic explanation and mathematical explanation », *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 44, 4, Part A, p. 477–492.

BROADBENT A., 2009, « Causation and models of disease in epidemiology », *Studies in history and philosophy of biological and biomedical sciences*, vol. 40, no. 4, p. 302–311.

—, 2013, *Philosophy of epidemiology*, New York, Palgrave Macmillan.

BROWN W. M., 1985, « On Defining "Disease" », *Journal of Medicine and Philosophy*, vol. 10, no. 4, p. 311–328.

BRUGGEMAN F. J. et WESTERHOFF H. V., 2007, « The nature of systems biology », *Trends in Microbiology*, vol. 15, no. 1, p. 45–50.

BRUTON O. C., 1952, « Agammaglobulinemia », *Pediatrics*, vol. 9, no. 6, p. 722–728.

—, 1952, « Absence of serum gamma globulins », *A.M.A. American Journal of Diseases of Children*, vol. 84, no. 5, p. 632–636.

BURIAN R. M., 1985, « On conceptual change in biology: The case of the gene », *Evolution at a Crossroads: The New Biology and the New Philosophy of Science*, D. Depew et B. Webster éds., Cambridge, MA, MIT Press, p. 21–42.

—, 2013, « Qu'est-ce qu'un gène : la fin d'une évidence ? », *La vie, et alors ? Débats passionnés de l'Antiquité à nos jours*, J. J. Kupiec ed., Paris, Belin, p. 115–130.

BURIAN R.M., 2004, « Molecular Epigenesis, Molecular Pleiotropy, and Molecular Gene Definitions », *History and Philosophy of the Life Sciences*, vol. 26, no. 1, p. 59–80.

BURLEY J. et HARRIS J., 2002, *A Companion to Genethics*, Malden, MA, Blackwell Publishers (Blackwell companions to philosophy), n° 21.

BURTON P. R. *et al., 2007,* « Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls »*, Nature*, vol. 447, no. 7145, p. 661–678.

BYNUM W. F., 1983, « Darwin and the doctors: evolution, diathesis, and germs in 19th-century Britain », *Gesnerus*, vol. 40, no. 1-2, p. 43–53.

—, 2002, « The evolution of germs and the evolution of disease: some British debates, 1870-1900 », *History and philosophy of the life sciences*, vol. 24, no. 1, p. 53–68.

CAMPANER R., 2011, « Understanding mechanisms in the health sciences », *Theoretical Medicine and Bioethics*, vol. 32, no. 1, p. 5–17.

CAPLAN A. L., 1992, « If gene therapy is the cure, what is the disease? », *Gene mapping: using laws and ethics as guides*, G. J. Annas et E. Sherman éds., New York, Oxford University Press, p. 128–141.

CAPOBIANCO Enrico, 2012, « Ten Challenges for Systems Medicine », *Frontiers in Genetics*, vol. 3, no. 193, p. 1–4.

CARTER K. C., 2003, *The rise of causal concepts of disease: case histories*, Burlington, VT, Ashgate (The history of medicine in context).

CARVUNIS A. R. *et al.,* 2009, « Biologie systémique: Des concepts d'hier aux découvertes de demain »*, M/S : médecine/sciences*, vol. 25, no. 6-7, p. 578–584.

CASADEVALL A. et PIROFSKI L. A., 1999, « Host-Pathogen Interactions: Redefining the Basic Concepts of Virulence and Pathogenicity », *Infection and Immunity*, vol. 67, no. 8, p. 3703–3713.

—, 2000, « Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease », *Infection and immunity*, vol. 68, no. 12, p. 6511–6518.

—, 2003, « The damage-response framework of microbial pathogenesis », *Nature Reviews Microbiology*, vol. 1, no. 1, p. 17–24.

CASANOVA J. L., 2000, « Mendelian predisposition to mycobacterial infections in humans », *Journal de la Société de Biologie*, vol. 194, no. 1, p. 25–28.

—, 2001, « Mendelian susceptibility to mycobacterial infection in man », *Swiss medical weekly*, vol. 131, no. 31/32, p. 445–454.

CASANOVA J. L. et ABEL L., 2004, « The human model: a genetic dissection of immunity to infection in natural conditions », *Nature Reviews Immunology*, vol. 4, no. 1, p. 55–66.

—, 2005, « Inborn errors of immunity to infection the rule rather than the exception », *The Journal of Experimental Medicine*, vol. 202, no. 2, p. 197–201.

—, 2007a, « Human genetics of infectious diseases: a unified theory », *The EMBO Journal*, vol. 26, no. 4, p. 915–922.

—, 2007b, « Primary immunodeficiencies: a field in its infancy », *Science (New York, N.Y.)*, vol. 317, no. 5838, p. 617–619.

—, 2009, « Revisiting Crohn's disease as a primary immunodeficiency of macrophages », *The Journal of Experimental Medicine*, vol. 206, no. 9, p. 1839–1843.

—, 2013, « The Genetic Theory of Infectious Diseases: A Brief History and Selected Illustrations », *Annual Review of Genomics and Human Genetics*, vol. 14, no. 1, p. 215–243.

CASANOVA J. L. *et al.,* 1996, « Idiopathic disseminated bacillus Calmette-Guérin infection: a French national retrospective study »*, Pediatrics*, vol. 98, 4, p. 774–778.

CASANOVA Jean-Laurent *et al.,* 2005, « From idiopathic infectious diseases to novel primary immunodeficiencies »*, Journal of Allergy and Clinical Immunology*, vol. 116, no. 2, p. 426–430.

CASPI A. *et al.,* 2002, « Role of genotype in the cycle of violence in maltreated children »*, Science*, vol. 297, no. 5582, p. 851–854.

CASROUGE A. *et al.,* 2006, « Herpes Simplex Virus Encephalitis in Human UNC-93B Deficiency »*, Science*, vol. 314, no. 5797, p. 308–312.

CASTRO O. *et al.,* 1994, « The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease »*, Blood*, vol. 84, no. 2, p. 643–649.

CAVALLI-SFORZA L. L. et BODMER W. F., 1971, *The Genetics of Human Populations*, San Francisco, W.H. Freeman and Company.

CHAN S. Y. et LOSCALZO J., 2012, « The Emerging Paradigm of Network Medicine in the Study of Human Disease », *Circulation research*, vol. 111, no. 3, p. 359–374.

DE CHASSEY B. *et al.,* 2008, « Hepatitis C virus infection protein network »*, Molecular Systems Biology*, vol. 4, no. 230, p. 1–12.

CHILDS B. et VALLE D., 2000, « Genetics, biology and disease », *Annual Review of Genomics and Human Genetics*, vol. 1, no. 1, p. 1–19.

CHOI M. *et al.,* 2009, « Genetic diagnosis by whole exome capture and massively parallel DNA sequencing »*, Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 45, p. 19096–19101.

CLERMONT G. *et al.,* 2009, « Bridging the gap between systems biology and medicine »*, Genome Medicine*, vol. 1, no. 9, p. 88.

COHEN M. L., 2000, « Changing patterns of infectious disease », *Nature*, vol. 406, no. 6797, p. 762–767.

COLLEGE NATIONAL DES ENSEIGNANTS ET PRATICIENS DE GENETIQUE MEDICALE, 2004, *Génétique médicale : formelle, chromosomique, moléculaire, clinique*, Paris, Masson.

COLLINS F. S., 2008, « Victor A. McKusick (1921-2008) », *Science*, vol. 321, no. 5891, p. 925–925.

CORDELL H. J., 2002, « Epistatis: what it means, what it doesn't mean and statistical methods to detect it in humans », *Human Molecular Genetics*, vol. 11, no. 20, p. 2463–2468.

—, 2009, « Detecting gene-gene interactions that underlie human diseases », *Nature Reviews Genetics*, vol. 10, , p. 392–404.

CORDELL H. J. et CLAYTON D.G., 2005, « Genetic association studies », *The Lancet*, vol. 366, no. 9491, p. 1121–1131.

COVENTRY P. A. et PICKSTONE J. V., 1999, « From what and why did genetics emerge as a medical specialism in the 1970s in the UK? A case-history of research, policy and services in the Manchester region of the NHS », *Social Science & Medicine*, vol. 49, no. 9, p. 1227–1238.

CRADDOCK N. *et al.,* 2010, « Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls »*, Nature*, vol. 464, no. 7289, p. 713–720.

D'AMICO R., 1995, « Is Disease a Natural Kind? », *Journal of Medicine and Philosophy*, vol. 20, no. 5, p. 551–569.

DANCHIN E. *et al.,* 2011, « Beyond DNA: integrating inclusive inheritance into an extended theory of evolution »*, Nature Reviews Genetics*, vol. 12, no. 7, p. 475–486.

DAWKINS R., 1976, *The Selfish Gene*, Oxford ; New York, Oxford University Press.

—, 1982, *The extended phenotype: the long reach of the gene*, Oxford ; New York, Oxford University Press.

DEAN M. *et al.,* 1996, « Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele fof the CCR5 structural gene »*, Science*, no. 273, p. 1856–1862.

DEBRET G. *et al.,* 2011, « Genetic susceptibility to a complex disease: the key role of functional redundancy »*, History and philosophy of the life sciences*, vol. 33, no. 4, p. 497–514.

DEKEUWER C., 2006, *Liberté de choix et destins individuels. Examen des concepts et des problèmes éthiques impliqués par la médecine prédictive*, Doctorat de philosophie, Université Panthéon-Sorbonne, Paris, France.

DEMPFLE A. *et al.,* 2008, « Gene-environment interactions for complex traits: definitions, methodological requirements and challenges »*, European Journal of Human Genetics*, vol. 16, no. 10, p. 1164–1172.

DEUTSCH J., 2012, *Le gène: un concept en évolution*, Paris, Éditions du Seuil.

DEWAN A., 2010, « Five Classic Articles in Genetic Epidemiology », *The Yale Journal of Biology and Medicine*, vol. 83, no. 2, p. 87–90.

—, 2006, « HTRA1 promoter polymorphism in wet age-related macular degeneration », *Science*, vol. 314, no. 5801, p. 989–992.

DICKSON S. P. *et al., 2010, « Rare Variants Create Synthetic Genome-Wide Associations », PLoS Biol*, vol. 8, no. 1, p. 1–12.

DIPPLE K. M. et MCCABE E. R. B., 2000a, « Modifier Genes Convert "Simple" Mendelian Disorders to Complex Traits », *Molecular Genetics and Metabolism*, vol. 71, no. 1-2, p. 43–50.

—, 2000b, « Phenotypes of patients with" simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. », *American Journal of Human Genetics*, vol. 66, no. 6, p. 1729–1735.

DO R., KATHIRESAN S. et ABECASIS G. R., 2012, « Exome sequencing and complex disease: practical aspects of rare variant association studies », *Human Molecular Genetics*, vol. 21, no. R1, p. R1–R9.

DOBZHANSKY T., 1973, « Nothing in biology makes sense except in the light of evolution », *American Biology Teacher*, vol. 35, no. 3, p. 125–129.

DUNCAN C. J. et SCOTT S., 2005, « What caused the Black Death? », *Postgraduate Medical Journal*, vol. 81, no. 955, p. 315–320.

DUPRÉ J., 2011, « Emerging sciences and new conceptions of disease; or, beyond the monogenomic differentiated cell lineage », *European Journal for Philosophy of Science*, vol. 1, no. 1, p. 119–131.

—, 2012, *Processes of Life: Essays in the Philosophy of Biology*, Oxford ; New York, Oxford University Press.

DUPRÉ J. et O'MALLEY M. A., 2007, « Metagenomics and biological ontology », *Studies in history and philosophy of biological and biomedical sciences*, vol. 38, no. 4, p. 834–846.

EBRAHIM S., 2006, « The future of modern epidemiology: genetics, methods, and history », *International Journal of Epidemiology*, vol. 35, no. 3, p. 511–512.

EDWARDS J. H., 1969, « Familial predisposition in man », *British Medical Bulletin*, no. 25, p. 58–64.

EICHLER E. E. *et al.,* 2010, « Missing heritability and strategies for finding the underlying causes of complex disease »*, Nature Reviews Genetics*, vol. 11, no. 6, p. 446–450.

ENCODE PROJECT CONSORTIUM *et al.,* 2012, « An integrated encyclopedia of DNA elements in the human genome »*, Nature*, vol. 489, no. 7414, p. 57–74.

ENGELHARDT H. T., 1975, « The Concepts of Health and Disease », *Evaluation and Explanation in the Biomedical Sciences*, H. T. Engelhardt et S. F. Spicker éds., Dordrecht, Holland ; Boston, D. Reidel Publishing Company (Philosophy and medicine ; v. 1), p. 125–141.

—, 1984, « Clinical Problems and the Concept of Disease », *Health, Disease, and Causal Explanations in Medicine*, L. Nordenfelt et B. I. B. Lindahl éds., Dordrecht, Boston, Lancaster, D. Reidel Publishing Company (Philosophy and medicine ; v. 16), p. 27–41.

—, 1996, *The Foundations of Bioethics*, Oxford ; New York, Oxford University Press.

ERESHEFSKY M., 2009, « Defining "Health" and "Disease" », *Studies in History and Philosophy of Science Part C*, vol. 40, no. 3, p. 221–227.

ESTELLER M., 2011, « Non-coding RNAs in human disease », *Nature Reviews Genetics*, vol. 12, no. 12, p. 861–874.

EVANS A. S., 1976, « Causation and disease: the Henle-Koch postulates revisited », *The Yale Journal of Biology and Medicine*, vol. 49, no. 2, p. 175–195.

—, 1991, « Causation and disease: effect of technology on postulates of causation. », *The Yale Journal of Biology and Medicine*, vol. 64, no. 5, p. 513–528.

FALK R., 2000, « The gene - a concept in tension », *The concept of the gene in development and evolution: historical and epistemological perspectives*, P. J. Beurton, R. Falk et H. J. Rheinberger éds., Cambridge ; New York, Cambridge University Press (Cambridge studies in philosophy and biology), p. 317–348.

FALKOW S., 1988, « Molecular Koch's postulates applied to microbial pathogenicity », *Reviews of infectious diseases*, vol. 10, Suppl 2, p. S274–276.

FANG F. C. et CASADEVALL A., 2011, « Reductionistic and Holistic Science », *Infection and Immunity*, vol. 79, no. 4, p. 1401–1404.

FEINBERG A. P., 2007, « Phenotypic plasticity and the epigenetics of human disease », *Nature*, vol. 447, no. 7143, p. 433–440.

FEINGOLD J., 2005, « Maladies multifactorielles : un cauchemar pour le généticien », *M/S : médecine sciences*, vol. 21, no. 11, p. 927–933.

FERGUSON-SMITH A. C., 2011, « Genomic imprinting: the emergence of an epigenetic paradigm », *Nature Reviews Genetics*, vol. 12, no. 8, p. 565–575.

FISHER R.A., 1918, « The correlation between relatives on the supposition of Mendelian Inheritance », *Transactions of the Royal Society of Edinburgh*, vol. 52, , p. 399–433.

FOX KELLER E., 2000, *The Century of the Gene*, Cambridge, Masachussets, Harvard University Press

—, 2005, « Revisiting “scale-free” networks », *BioEssays*, vol. 27, no. 10, p. 1060–1068.

VAN FRAASSEN B. C., 1980, *The Scientific Image*, Oxford ; New York, Oxford University Press.

FRIGG R. et HARTMANN S., 2006, « Models in Science », *The Stanford Encyclopedia of Philosophy*, E. N. Zalta ed., Stanford.

GALAS D. J. et HOOD Leroy, 2009, « Systems Biology and Emerging Technologies Will Catalyze the Transition from Reactive Medicine to Predictive, Personalized, Preventive and Participatory (P4) Medicine », *Interdisciplinary Bio Central*, vol. 1, no. 2, p. 1–4.

GANNETT L., 1999, « What’s in a cause?: the pragmatic dimensions of genetic explanations », *Biology and Philosophy*, vol. 14, no. 3, p. 349–373.

GARROD A., 1931, *Inborn Factors of Disease*, London, UK, Clarendon.

GASCHIGNARD J., SCURR E. et ALCAÏS A., 2013, « La lèpre, pilier de la génétique des maladies infectieuses », *Pathologie Biologie*, vol. 61, no. 3, p. 120–128.

GAUDILLIERE J. P., 2000, « Mendelism and medicine: controlling human inheritance in local contexts, 1920–1960 », *Comptes Rendus de l’Académie des Sciences-Series III-Sciences de la Vie*, vol. 323, no. 12, p. 1117–1126.

GAYON J., 1992, *Darwin et l’après-Darwin. Une histoire de l’hypothèse de sélection naturelle*, Paris, Kimé.

—, 1993, « La biologie entre loi et histoire », *Philosophie*, vol. 38, , p. 30–57.

—, 1997, « The paramount power of selection: from Darwin to Kauffman », *Structures and Norms in Science, Volume Two of the Tenth International Congress of Logic, Methodology and Philosophy of Science*, M. L. Dalla Chiara *et al.* éds., Dordrecht, Pays-Bas, Kluwer, p. 265–282.

—, 2000, « From measurement to organization: a philosophical scheme for the history of the concept of heredity », The concept of the gene in development and evolution: historical and epistemological perspectives, P. J. Beurton, R. Falk et H. J. Rheinberger éds., Cambridge ; New York, Cambridge University Press (Cambridge studies in philosophy and biology), p. 69–90.

—, 2009a, « Déterminisme génétique, déterminisme bernardien, déterminisme laplacien », *Le hasard au cœur de la cellule: probabilités, déterminisme,*

génétique, J. J. Kupiec éd., Éditions Matériologiques, Paris, Syllepse, p. 115–129.

—, 2009b, « Mort ou persistance du darwinisme ? Regard d'un épistémologue », *Comptes Rendues Palevol*, vol. 8, , p. 321–340.

GAYON J. et BURIAN R. M., 2004, « National traditions and the emergence of genetics: the French example », *Nature Reviews Genetics*, vol. 5, no. 2, p. 150–156.

GAYON J., JACOBI D. et LORNE M. C., 2006, *L'éternel retour de l'eugénisme*, Paris, Presses Universitaires de France.

GAYON Jean, 2007, « The Concept of the Gene in Contemporary Biology: Continuity or Dissolution? », *The Influence of Genetics on Contemporary Thinking*, A. Fagot-Largeault, S. Rahman et J. M. Torres éds., Springer Netherlands, n° 6, p. 81–95.

GERMAIN P. L., RATTI E. et BOEM F., 2014, « Junk or functional DNA? ENCODE and the function controversy », *Biology & Philosophy*, vol. 29, no. 3, p. 1–25.

GIBSON G., 2009, « Decanalization and the origin of complex disease », *Nature Reviews Genetics*, vol. 10, no. 2, p. 134–140.

—, 2012, « Rare and common variants: twenty arguments », *Nature Reviews Genetics*, vol. 13, no. 2, p. 135–145.

GIERE R. N., 2004, « How Models Are Used to Represent Reality », *Philosophy of Science*, vol. 71, no. 5, p. 742–752.

GIFFORD F., 1989, « Complex genetic causation of human disease: Critiques of and rationales for heritability and path analysis », *Theoretical Medicine and Bioethics*, vol. 10, no. 2, p. 107–122.

—, 1990, « Genetic traits », *Biology and Philosophy*, vol. 5, no. 3, p. 327–347.

GIROUX E., 2010a, « Philosophie de la médecine », *Précis de philosophie des sciences*, D. Bonnay et M. Cozick éds., Paris, Vuibert, p. 404–441.

—, 2010b, *Après Canguilhem: Définir la santé et la maladie*, Paris, Presses Universitaires de France.

GIROUX E. et LEMOINE M., 2012, *Philosophie de la médecine. Volume 2 : Santé, maladie, pathologie*, Paris, Vrin.

GLENNAN S., 2005, « Modeling mechanisms », *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 36, no. 2, p. 443–464.

GODFREY-SMITH P., 1996, *Complexity and the function of mind in nature*, Cambridge, Cambridge University Press.

—, 2000, « The Replicator in Retrospect », *Biology and Philosophy*, vol. 15, no. 3, p. 403–423.

—, 2009, *Darwinian populations and natural selection*, Oxford ; New York, Oxford University Press.

GOH G. et CHOI M., 2012, « Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases », *Genomics & Informatics*, vol. 10, no. 4, p. 214–219.

GOH K. I. et CHOI I. G., 2012, « Exploring the human diseasome: the human disease network », *Briefings in Functional Genomics*, vol. 11, no. 6, p. 533–542.

GOH K. I. *et al.,* 2007, « The human disease network »*, Proceedings of the National Academy of Sciences*, vol. 104, no. 21, p. 8685–8690.

GOULD S. J., 1981, *The mismeasure of man*, First ed., New York, Norton.

GOULD S. J. et LEWONTIN R. C., 1979, « The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme », *Proceedings of the Royal Society of London B*, vol. 205, no. 1161, p. 581–598.

GREEN M. J. et BOTKIN J. R., 2003, « "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests », *Annals of internal medicine*, vol. 138, no. 7, p. 571–575.

GRIFFITHS P. E., 2001, « Genetic Information: A Metaphor in Search of a Theory », *Philosophy of Science*, vol. 68, no. 3, p. 394–412.

GRIFFITHS P. E. et GRAY R. D., 1994, « Developmental Systems and Evolutionary Explanation », *The Journal of Philosophy*, vol. 91, no. 6, p. 277–304.

—, 2001, « Darwinism and Developmental Systems », *Cycles of contingency*, S. Oyama, P. E. Griffiths et R. D. Gray éds., MIT Press., Cambridge, MA, p. 195–218.

—, 2005, « Discussion: Three Ways to Misunderstand Developmental Systems Theory », *Biology & Philosophy*, vol. 20, no. 2-3, p. 417–425.

GRIFFITHS P. E. et KNIGHT R. D., 1998, « What is the developmentalist challenge? », *Philosophy of Science*, vol. 65, no. 2, p. 253–258.

GRIFFITHS P. E. et NEUMANN-HELD E. M., 1999, « The many faces of the gene », *BioScience*, vol. 49, no. 8, p. 656–662.

GRIFFITHS P. E. et STOTZ K., 2006, « Genes in the postgenomic era », *Theoretical medicine and bioethics*, vol. 27, no. 6, p. 499–521.

GRIFFITHS P. E. et TABERY J., 2008, « Behavioral genetics and development: Historical and conceptual causes of controversy », *New Ideas in Psychology*, vol. 26, no. 3, p. 332–352.

GRMEK M. D., 1993, « Le Concept de Maladie Émergente », *History and Philosophy of the Life Sciences*, vol. 15, no. 3, p. 281–296.

GUO Y. *et al.,* 2011, « Herpes simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is otherwise redundant in protective immunity »*, The Journal of Experimental Medicine*, vol. 208, no. 10, p. 2083–2098.

GUTTMACHER A. E. et COLLINS F. S., 2002, « Genomic Medicine — A Primer », *New England Journal of Medicine*, vol. 347, no. 19, p. 1512–1520.

HABERSTICK B. C. *et al.,* 2005, « Monoamine oxidase A (MAOA) and antisocial behaviors in the presence of childhood and adolescent maltreatment »*, American Journal of Medical Genetics*, vol. 135, no. 1, p. 59–64.

HACEIN-BEY-ABINA S. *et al.,* 2010, « Efficacy of Gene Therapy for X-Linked Severe Combined Immunodeficiency »*, The New England Journal of Medicine*, vol. 363, no. 4, p. 355–64.

HACKING I., 2007, « Natural Kinds: Rosy Dawn, Scholastic Twilight », *Royal Institute of Philosophy Supplements*, vol. 82, no. 61, p. 203–239.

HAGEN J. B., 2000, « The origins of bioinformatics », *Nature Reviews Genetics*, vol. 1, no. 3, p. 231–236.

HAMOSH A. *et al.,* 2005, « Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders »*, Nucleic Acids Research*, vol. 33, Database Issue, p. D514–D517.

HAN J. D. J. *et al.,* 2004, « Evidence for dynamically organized modularity in the yeast protein–protein interaction network »*, Nature*, vol. 430, no. 6995, p. 88–93.

HANAHAN D. et WEINBERG R. A., 2000, « The hallmarks of cancer », *Cell*, vol. 100, no. 1, p. 57–70.

—, 2011, « Hallmarks of cancer: the next generation », *Cell*, vol. 144, no. 5, p. 646–674.

HARPER P. S., 2004, *Landmarks in medical genetics: Classic papers with commentaries*, Oxford ; New York, Oxford University Press (Oxford monographs on medical genetics), n° 51

—, 2008, *A short history of medical genetics*, Oxford ; New York, Oxford University Press (Oxford monographs on medical genetics), n° 57.

—, 2012, « Victor McKusick and the History of Medical Genetics », *Victor McKusick and the History of Medical Genetics*, K. R. Dronamraju et C. Francomano éds., New York, Springer, p. 145–161.

HEDGECOE A. M., 1998, « Geneticization, medicalisation and polemics », *Medicine, Health Care and Philosophy*, vol. 1, no. 3, p. 235–243.

—, 2003, « Expansion and uncertainty: cystic fibrosis, classification and genetics », *Sociology of Health & Illness*, vol. 25, no. 1, p. 50–70.

HEDGECOE Alan M., 2002, « Reinventing diabetes: classification, division and the geneticization of disease », *New Genetics and Society*, vol. 21, no. 1, p. 7–27.

HEMPEL C. G., 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York, Free Press.

HERMAN M. *et al.,* 2012, « Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes simplex encephalitis of childhood *», Journal of Experimental Medicine*, vol. 209, no. 9, p. 1567–1582.

AL-HERZ W. *et al.,* 2011, « Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency *», Frontiers in immunology*, vol. 2, no. 54, p. 1–26.

HESSLOW G., 1984, « What is a genetic disease? », *Health, disease, and causal explanations in medicine*, L. Nordenfelt et B. I. B. Lindahl éds., Dordrecht, Pays-Bas, D. Reidel publ. Co (Philosophy and medicine ; v. 16), p. 183–193.

—, 1988, « The Problem of Causal Selection », *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality*, D. J. Hilton ed., Brighton, Harverster Press, p. 11–31.

HIRSCHHORN J. N. et DALY M. J., 2005, « Genome-wide association studies for common diseases and complex traits », *Nature Reviews Genetics*, vol. 6, no. 2, p. 95–108.

HOEDEMAEKERS R. et TEN HAVE H., 1997, « Geneticisation: The cyprus paradigm », *The Journal of Medicine and Philosophy*, vol. 23, no. 3, p. 274–287.

HOFMANN B., 2001, « Complexity of the concept of disease as shown through rival theoretical frameworks », *Theoretical medicine and bioethics*, vol. 22, no. 3, p. 211–236.

HOLM S., 1999, « There is Nothing Special about Genetic Information », *Genetic Information: Acquisition, Access, and Control*, A. K. Thompson et R. F. Chadwick éds., New York, Kluwer Acadamic / Plenum Publishers, p. 97–103.

HOLTZMAN N. A., 2001, « Putting the search for genes in perspective », *International journal of health services: planning, administration, evaluation*, vol. 31, no. 2, p. 445–461.

HOOD L., 1998, « Systems biology: new opportunities arising from genomics, proteomics and beyond », *Experimental Hematology*, vol. 26, , p. 681–681.

—, 2004, « Systems Biology and New Technologies Enable Predictive and Preventative Medicine », *Science*, vol. 306, no. 5696, p. 640–643.

HOOD L. et FLORES M., 2012, « A Personal View on Systems Medicine and the Emergence of Proactive P4 Medicine: Predictive, Preventive, Personalized and Participatory », *New Biotechnology*, vol. 29, no. 6, p. 613–624.

HOOPER L. V., LITTMAN D. R. et MACPHERSON A. J., 2012, « Interactions Between the Microbiota and the Immune System », *Science*, vol. 336, no. 6086, p. 1268–1273.

HULL D. L., 1980, « Individuality and Selection », *Annual Review of Ecology and Systematics*, vol. 11, no. 1, p. 311–332.

HULL R. T., 1979a, « Why "Genetic Disease"? », *Genetic counseling: facts, values, and norms*, A. M. Capron *et al. éds., New York, A. R. Liss (Birth defects original article series).*

—, 1979b, « On Getting "Genetic" Out of "Genetic Disease" », Contemporary Issues in Biomedical Ethics, J. W. Davis, B. Hoffmaster et S. Shorten éds., Clifton, N. J., Humana Press, p. 71–87.

—, 1997, « Genetic carrier status and the contextual reality of genetic disease: a contribution to URAM genome studies », *Ultimate Reality and Meaning: Interdisciplinary Studies in the Philosophy of Understanding*, vol. 20, no. 4, p. 304–316.

HUNEMAN P., 2010, « Topological explanations and robustness in biological sciences », *Synthese*, vol. 177, no. 2, p. 213–245.

—, 2013, « Causal parity and Externalisms: Extensions in Life and Mind », *Minds and Machines*, vol. 23, no. 3, p. 1–28.

HUNTER D. J., 2005, « Gene-environment interactions in human diseases », *Nature Reviews Genetics*, vol. 6, no. 4, p. 287–298.

VAN IJZENDOORN M. H. *et al.,* 2011, « Gene-by-environment experiments: a new approach to finding the missing heritability »*, Nature Reviews Genetics*, vol. 12, no. 12, p. 881–881.

ILKILIC I., 2009, « Coming to Grips with Genetic Exceptionalism: Roots and Reach of an Explanatory Model », *Medicine Studies*, vol. 1, no. 2, p. 131–142.

JABLONKA E. et LAMB M. J., 2002, « The changing concept of epigenetics », *Annals of the New York Academy of Sciences*, vol. 981, no. 1, p. 82–96.

JACUNSKI A. et TATONETTI N. P., 2013, « Connecting the dots: applications of network medicine in pharmacology and disease », *Clinical pharmacology and therapeutics*, vol. 94, no. 6, p. 659–669.

JANJIC V. et PRZULJ N., 2012, « Biological function through network topology: a survey of the human diseasome », *Briefings in Functional Genomics*, vol. 11, no. 6, p. 522–532.

JENSEN A. R., 1969, « How much can we boost IQ and scholastic achievement? », *Harvard Educational Review*, vol. 39, no. 1, p. 1–123.

JEONG H. *et al.,* 2001, « Lethality and centrality in protein networks »*, Nature*, vol. 411, no. 6833, p. 41–42.

—, 2000, « The large-scale organization of metabolic networks », *Nature*, vol. 407, no. 5, p. 651–652.

JIANG Y. H., BRESSLER J. et BEAUDET A. L., 2004, « Epigenetics and Human Disease », *Annual Review of Genomics and Human Genetics*, vol. 5, , p. 479–510.

JIMENEZ-SANCHEZ G., CHILDS B. et VALLE D., 2001, « Human disease genes », *Nature*, vol. 409, no. 6822, p. 853–855.

JORDAN B., 1988, « Grandeur et servitudes de la génétique inverse », *M/S : médecine/ sciences*, vol. 4, no. 3, p. 138–140.

—, 2006, « La “chasse aux gènes”: méthodes, enjeux et conséquences », *L’éternel retour de l’eugénisme*, J. Gayon et D. Jacobi éds., Paris, Presses Universitaires de France, p. 49–59.

—, 2010, « A la recherche de l’héritabilité perdue », *M/S : médecine/sciences*, vol. 26, no. 5, p. 541–543.

—, 2012, « Rare is frequent », *M/S : médecine/sciences*, vol. 28, no. 10, p. 893–896.

KAKUK P., 2008, « Gene Concepts and Genethics: Beyond Exceptionalism », *Science and Engineering Ethics*, vol. 14, no. 3, p. 357–375.

KAVALOSKI V., 1974, *The Vera Causa Principle: A Historico-Philosophical Study of a Metatheoretical Concept from Newton through Darwin*, Thèse de doctorat, University of Chicago, Chicago.

KENDLER K. S., ZACHAR P. et CRAVER C., 2011, « What kinds of things are psychiatric disorders? », *Psychological medicine*, vol. 41, no. 6, p. 1143–1150.

KERR A., 2000, « (Re)Constructing Genetic Disease: The Clinical Continuum between Cystic Fibrosis and Male Infertility », *Social Studies of Science*, vol. 30, no. 6, p. 847–894.

—, 2004, « Giving up on geneticization: a comment on Hedgecoe’s “Expansion and uncertainty: cystic fibrosis, classification and genetics” », *Sociology of health & illness*, vol. 26, no. 1, p. 102–106.

—, 2005, « Understanding genetic disease in a socio-historical context: a case study of cystic fibrosis », *Sociology of Health & Illness*, vol. 27, no. 7, p. 873–896.

KERR Anne, 2005, « Understanding genetic disease in a socio-historical context: a case study of cystic fibrosis », *Sociology of Health & Illness*, vol. 27, no. 7, p. 873–896.

KEVLES D. J. et HOOD L., 1992, *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, Harvard, Harvard University Press.

KHOURY M. J., 1997, « Genetic Epidemiology and the Future of Disease Prevention and Public Health », *Epidemiologic Reviews*, vol. 19, no. 1, p. 175–180.

—, 2003, « Genetics and genomics in practice: the continuum from genetic disease to genetic information in health and disease », *Genetics in Medicine*, vol. 5, no. 4, p. 261–268.

KHOURY M. J., BEATY T. H. et COHEN B. H., 1993, *Fundamentals of Genetic Epidemiology*, Oxford, Oxford University Press.

KHOURY M. J. *et al.,* 2004, « The emergence of epidemiology in the genomics age », *International Journal of Epidemiology*, vol. 33, no. 5, p. 936–944.

KIM-COHEN J. *et al.,* 2006, « MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis », *Molecular psychiatry*, vol. 11, no. 10, p. 903–913.

KINCAID H., 2008, « Do We Need Theory to Study Disease?: lessons from cancer research and their implications for mental illness », *Perspectives in Biology and Medicine*, vol. 51, no. 3, p. 367–378.

KITANO H., 2002, « Systems Biology: A Brief Overview », *Science*, vol. 295, no. 5560, p. 1662–1664.

—, 2004, « Biological robustness », *Nature Reviews Genetics*, vol. 5, no. 11, p. 826–837.

KITCHER P., 2001, « Battling the undead: how and how not to resist genetic determinism », *Thinking about evolution: historical, philosophical and political perspectives*, R. S. Singh *et al. éds., Cambridge University Press., Cambridge, p. 396–414.*

KLEIN Christine, 2012, « The Promise and Limitations of Genome-wide Association Studies », JAMA: The Journal of the American Medical Association, vol. 308, no. 18, p. 1867–1868.

KLEIN R. J. *et al., 2005, « Complement Factor H Polymorphism in Age-Related Macular Degeneration », Science*, vol. 308, no. 5720, p. 385–389.

KNUDSON A. G., 1971, « Mutation and cancer: statistical study of retinoblastoma », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, no. 4, p. 820–823.

KOLCH W. et KHOLODENKO B., 2013, « Systems medicine: opportunities and challenges for systems biology approaches », *FEBS Journal*, vol. 280, no. 23, p. 5937–5937.

KU C. S. *et al.,* 2010, « The pursuit of genome-wide association studies: where are we now? »*, Journal of Human Genetics*, vol. 55, no. 4, p. 195–206.

KUHN T. S., 1962, *The structure of scientific revolutions*, Chicago, IL, US, University of Chicago Press.

LAKATOS I., 1976, « Falsification and the Methodology of Scientific Research Programmes », *Can Theories be Refuted?*, S. G. Harding ed., Dordrecht, Springer Netherlands (Synthese Library), n˚ 81, p. 205–259.

—, 1980, *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*, Cambridge, Cambridge University Press.

LALOUEL J. M. *et al.,* 1983, « A unified model for complex segregation analysis »*, American Journal of Human Genetics*, no. 35, p. 816–826.

LANDER E. S., 1996, « The New Genomics: Global Views of Biology », *Science*, vol. 274, no. 5287, p. 536–539.

LANGE M., 2007, « The End of Diseases », *Philosophical Topics*, vol. 35, no. 1, p. 265–292.

LAPLANE L., 2013, *Cellules souches cancéreuses : ontologie et thérapies*, Thèse de doctorat, Université Paris Ouest Nanterre, Paris.

—, 2014, « Identifying theories in developmental biology : the case of the cancer stem cell theory », *Towards a Theory of Development*, Oxford ; New York, Oxford University Publishing.

LAWRENCE T. *et al.,* 2005, « Autosomal-dominant primary immunodeficiencies »*, Current opinion in hematology*, vol. 12, no. 1, p. 22–30.

LEAVITT J. W., 1996, *Typhoid Mary: captive to the public's health*, Boston, Beacon Press.

LEDERBERG J., 2000, « Infectious History », *Science*, vol. 288, no. 5464, p. 287–293.

LEE P. et HUDSON T. J., 2000, « La puce à ADN en médecine et en science. », *M/S : médecine/sciences*, vol. 16, no. 1, p. 43.

LEE S. H. *et al.,* 2011, « Estimating Missing Heritability for Disease from Genome-wide Association Studies »*, The American Journal of Human Genetics*, vol. 88, no. 3, p. 294–305.

LEMOINE M., à venir, « Trois conceptions sémantiques des théories en médecine », *Lato Sensu*, p. accepté.

—, 2011, *La désunité de la médecine: essai sur les valeurs explicatives de la science médicale*, Paris, Hermann.

—, 2013a, « The naturalization of the concept of disease », *Classification, Disease and Evience. New Essays in the philosophy of medicine*, G. Lambert, M. Silberstein et P. Huneman éds., Dordrecht, Pays-Bas, Springer.

—, 2013b, « Defining disease beyond conceptual analysis: an analysis of conceptual analysis in philosophy of medicine », *Theoretical Medicine and Bioethics*, vol. 34, no. 4, p. 309–325.

LESNE A., 2008, « Robustness: confronting lessons from physics and biology », *Biological reviews of the Cambridge Philosophical Society*, vol. 83, no. 4, p. 509–532.

—, 2009, « Biologie des systèmes: L'organisation multiéchelle des systèmes vivants », *M/S : médecine/sciences*, vol. 25, no. 6-7, p. 585–587.

LEVINS R. et LEWONTIN R.C., 1985, *The Dialectical Biologist*, Cambridge, MA, Harvard University Press.

LEVY A. et BECHTEL W., 2013, « Abstraction and the Organization of Mechanisms », *Philosophy of Science*, vol. 80, no. 2, p. 241–261.

LEWENS T., 2004, « What is genethics? », *Journal of Medical Ethics*, vol. 30, no. 3, p. 326–328.

LEWONTIN R. C., 1974, « The analysis of variance and the analysis of causes », *American Journal of Human Genetics*, vol. 26, no. 3, p. 400–411.

—, 2000, *The Triple Helix: Gene, Organism and Environment*, Cambridge, MA, Harvard University Press.

LINDEE M. S., 2000, « Genetic disease since 1945 », *Nature Reviews Genetics*, vol. 1, no. 3, p. 236–241.

—, 2002, « Genetic disease in the 1960s: A structural revolution », *American Journal of Medical Genetics*, vol. 115, no. 2, p. 75–82.

LINDEE M. S. et MUELLER R., 2011, « Is Cystic Fibrosis Genetic Medicine's Canary? », *Perspectives in Biology and Medicine*, vol. 54, no. 3, p. 316–331.

LIPPMAN A., 1991, « Prenatal genetic testing and screening: constructing needs and reinforcing inequalities », *American Journal of Law and Medicine*, vol. 17, no. 1-2, p. 11–50.

—, 1994, « The genetic construction of prenatal testing: choice, consent or conformity for women », *Women and prenatal testing: facing the challenges of genetic testing*, K. H. Rosenberg et E. J. Thomson éds., Columbus, Ohio State University, p. 9–34.

LIU R *et al.,* 1996, « Homozygous defects in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection »*, Cell*, no. 86, p. 367–377.

LOSCALZO J., 2012, « Personalized Cardiovascular Medicine and Drug Development Time for a New Paradigm », *Circulation*, vol. 125, no. 4, p. 638–645.

LOSCALZO J., KOHANE I. et BARABASI A. L., 2007, « Human disease classification in the postgenomic era: a complex systems approach to human pathobiology », *Molecular systems biology*, vol. 3, no. 1, p. 1–11.

LOTTENBERG R. et HASSELL K. L., 2005, « An Evidence-Based Approach to the Treatment of Adults with Sickle Cell Disease », *ASH Education Program Book*, vol. 2005, no. 1, p. 58–65.

LUSH J. L., 1940, « Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics », *Proceedings of the American Society for Animal Production*, vol. 33, , p. 293–301.

MACHAMER P., 2004, « Activities and Causation: The Metaphysics and Epistemology of Mechanisms », *International Studies in the Philosophy of Science*, vol. 18, no. 1, p. 27–39.

MACHAMER P., DARDEN L. et CRAVER C. F., 2000, « Thinking about mechanisms », *Philosophy of Science*, vol. 67, no. 1, p. 1–25.

MACKIE J. L., 1965, « Causes and conditions », *American philosophical quarterly*, vol. 2, no. 4, p. 245–264.

MAGNUS D., 2004, « The concept of genetic disease », *Health, disease, and illness: concepts in medicine*, A. L. Caplan, J. J. McCartney et D. A. Sisti éds., Washington, D.C., Georgetown University Press, p. 233–242.

MANOLIO T. A. *et al., 2009, « Finding the missing heritability of complex diseases », Nature*, vol. 461, no. 7265, p. 747–753.

MAUBARET C. G. *et al.,* 2011, « Autosomal Dominant Retinitis Pigmentosa with Intrafamilial Variability and Incomplete Penetrance in Two Families Carrying Mutations in PRPF8 »*, Investigative Ophthalmology & Visual Science*, vol. 52, no. 13, p. 9304–9309.

MAYR E., 1961, « Cause and Effect in Biology », *Science, New Series*, vol. 134, no. 3489, p. 1501–1506.

MCCARTHY M. I. *et al.,* 2008, « Genome-wide association studies for complex traits: consensus, uncertainty and challenges »*, Nature Reviews Genetics*, vol. 9, no. 5, p. 356–369.

MCKUSICK V. A., 1969, « On lumpers and splitters, or the nosology of genetic disease », *Birth Defects: Original Article Series*, vol. 5, no. 1, p. 23–32.

—, 2007, « Mendelian Inheritance in Man and Its Online Version, OMIM », *American Journal of Human Genetics*, vol. 80, no. 4, p. 588–604.

MEANEY M. J., 2010, « Epigenetics and the biological definition of gene x environment interactions », *Child development*, vol. 81, no. 1, p. 41–79.

MEARS J. G. *et al., 1981, « The sickle gene polymorphism in North Africa », Blood*, vol. 58, no. 3, p. 599–601.

MELENDRO-OLIVER S., 2004, « Shifting Concepts of Genetic Disease », *Science Studies*, vol. 17, no. 1, p. 20–33.

MERLIN F., 2010, « On Griffiths and Gray's Concept of Expanded and Diffused Inheritance », *Biological Theory*, vol. 5, no. 3, p. 206–215.

METHOT P. O., 2009, « Darwin et la médecine : intérêt et limites des explications évolutionnaires en médecine », *Les mondes darwiniens: l'évolution de l'évolution*, T. Heams *et al.* éds., Paris, Syllepse, p. 658–684.

—, 2010, « Research traditions and evolutionary explanations in medicine », *Theoretical Medicine and Bioethics*, vol. 32, no. 1, p. 75–90.

—, 2012a, *Historical epistemology of the concept of virulence: molecular, ecological, and evolutionary perpectives on emerging infectious diseases in the 19th and 20th century*, Thèse de doctorat, University of Exeter, GB.

—, 2012b, « Why do parasites harm their host? On the origin and legacy of Theobald Smith's "law of declining virulence"-1900-1980 », *History and philosophy of the life sciences*, vol. 34, no. 4, p. 561–601.

—, 2012c, « Understanding pathogens in the era of next generation sequencing », *The Journal of Infection in Developing Countries*, vol. 6, no. 09.

MILL J. S., 1869, « On the composition of causes », *System of Logic, Rationative and Inductive*, New York, Harpers and Brothers, Publishers.

MILLER F. A. *et al.,* 2005, « Ruling in and ruling out: Implications of molecular genetic diagnoses for disease classification », *Social Science & Medicine*, vol. 61, no. 12, p. 2536–2545.

MILLER L. H. *et al.,* 1976, « The resistance factor to Plasmodium vivax in blacks. The duffy-blood-group genotype, FyFy », *New England Journal of Medicine*, vol. 295, no. 6, p. 302–304.

MITCHELL S. D., 2003, *Biological complexity and integrative pluralism*, Cambridge, UK ; New York, N.Y, Cambridge University Press (Cambridge studies in philosophy and biology).

—, 2006, « Modularity-More Than a Buzzword? », *Biological Theory*, vol. 1, no. 1, p. 98–101.

MITCHELL S. D. et DIETRICH M. R., 2006, « Integration without Unification: An Argument for Pluralism in the Biological Sciences », *The American Naturalist*, vol. 168, S6, p. S73–S79.

DE MONTALEMBERT M. *et al.,* 1993, « Epidemiological and clinical study of sickle cell disease in France, French Guiana and Algeria »*, European Journal of Haematology*, vol. 51, no. 3, p. 136–140.

MONTOYA J. M., PIMM S. L. et SOLÉ R. V., 2006, « Ecological networks and their fragility », *Nature*, vol. 442, no. 7100, p. 259–264.

MOORE J. H., 2003, « The ubiquitous nature of epistasis in determining susceptibility to common human disease », *Human Heredity*, vol. 56, no. 1-3, p. 73–82.

MOORE J. H. et WILLIAMS S. M., 2009, « Epistasis and its implications for personal genetics », *The American Journal of Human Genetics*, vol. 85, , p. 309–320.

MORANGE M., 2002, « The relations between genetics and epigenetics: a historical point of view », *Annals of the New York Academy of Sciences*, vol. 981, , p. 50–60.

—, 2005, « Quelle place pour l'épigénétique ? », *M/S : médecine sciences*, vol. 21, no. 4, p. 367–369.

—, 2011, « Les mirages de l'épigénétique », *Critique*, n° 764-765, no. 1, p. 153–158.

MORTON N. E., 2006, « Fifty years of genetic epidemiology, with special reference to Japan », *Journal of Human Genetics*, vol. 51, no. 4, p. 269–277.

MORTON N. E. et CHUNG C. S., 1978, *Genetic epidemiology*, New York, Academic Press.

MOULIN A-M., 1991, *Le dernier langage de la médecine - Histoire de l'immunologie de Pasteur au Sida*, Paris, Presses Universitaires de France.

AL-MUHSEN S. et CASANOVA J. L., 2008, « The genetic heterogeneity of mendelian susceptibility to mycobacterial diseases », *The Journal of allergy and clinical immunology*, vol. 122, no. 6, p. 1043–1051.

MUNSON R., 1981, « Why Medicine Cannot be a Science », *Journal of Medicine and Philosophy*, vol. 6, no. 2, p. 183–208.

MURPHY D., 2005, « The Concept of Mental Illness--Where the Debate has Reached and Where it Needs to Go », *Journal of Theoretical and Philosophical Psychology*, vol. 25, no. 1, p. 116–132.

—, 2010, « Explanation in Psychiatry: Explanation in Psychiatry », *Philosophy Compass*, vol. 5, no. 7, p. 602–610.

NACHER J. C. et SCHWARTZ J. M., 2012, « Modularity in protein complex and drug interactions reveals new polypharmacological properties », *PloS one*, vol. 7, no. 1, p. 1–13.

NAGEL E., 1961, *The structure of science: problems in the logic of scientific explanation*, London, Routledge & Kegan Paul.

NAVRATIL V. *et al.,* 2011, « When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases »*, BMC systems biology*, vol. 5, no. 13, p. 1–15.

—, 2009, « VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks », *Nucleic Acids Research*, vol. 37, Database, p. D661–D668.

NAVRATIL V., LOTTEAU V. et RABOURDIN-COMBE C., 2010, « La cellule infectée virtuelle », *M/S : médecine/sciences*, vol. 26, no. 6-7, p. 603–609.

NEEL J. V. et SCHULL W. J., 1954, *Human heredity*, Chicago, University of Chicago Press.

NELKIN D. et LINDEE M. S., 1995, *The DNA Mystique: The Gene as a Cultural Icon*, Ann Arbor, University of Michigan Press.

—, 1996, « "Genes made me do it": the appeal of biological explanations, Politics and the Life Sciences », *Politics and the Life Sciences*, vol. 15, no. 1, p. 95–96.

NELSON M. R. *et al.,* 2012, « An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People »*, Science*, vol. 337, no. 6090, p. 100–104.

NESSE R. M., 1994, *Why We Get Sick: The New Science of Darwinian Medicine*, New York, Times Books.

—, 2001, « How is Darwinian medicine useful? », *Western Journal of Medicine*, vol. 174, no. 5, p. 358–360.

—, 2008, « Evolution: medicine's most basic science », *The Lancet*, vol. 372, S21–S27.

—, 2010, « Making evolutionary biology a basic science for medicine », *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, Suppl 1, p. 1800–1807.

NESSE R. M. et STEARNS S. C., 2008, « The great opportunity: evolutionary applications to medicine and public health », *Evolutionary Applications*, vol. 1, no. 1, p. 28–48.

NICHOLSON J. K. *et al., 2012,* « Host-Gut Microbiota Metabolic Interactions »*, Science*, vol. 336, no. 6086, p. 1262–1267.

NORDENFELT L., 1995, *On the Nature of Health: An Action-Theoretic Approach*, Dordrecht, The Netherlands, Kluwer Academic Publishers.

—, 2001, *Health, Science, and Ordinary Language*, Amsterdam ; New York, Rodopi.

—, 2007, « The concepts of health and illness revisited », *Medicine, Health Care and Philosophy*, vol. 10, no. 1, p. 5–10.

OHNO S., 1970, *Evolution by gene duplication.*, New York, Springer-Verlag.

OMS WHO, 2006, *ICD-10. International Statistical Classification of Diseases and Related Health Problems. Tenth revision. Volume 2. Second Edition*, Genève, OMS/WHO.

ONNELA J. P. *et al.,* 2007, « Structure and tie strengths in mobile communication networks »*, Proceedings of the National Academy of Sciences*, vol. 104, no. 18, p. 7332–7336.

OTI M. et BRUNNER H. G., 2006, « The modular nature of genetic diseases », *Clinical genetics*, vol. 71, no. 1, p. 1–11.

OYAMA S., 1985, *The ontogeny of information*, Cambridge, Cambridge University Press.

—, 2000, « Causal democracy and causal contributions in developmental systems theory », *Philosophy of Science*, vol. 67, no. 3, p. S332–S347.

OYAMA S., GRIFFITHS P. E. et GRAY R. D., 2001, *Cycles of contingency: Developmental systems and evolution*, Cambridge, MA, MIT Press.

PARK J. *et al.,* 2009, « The impact of cellular networks on disease comorbidity », *Molecular Systems Biology*, vol. 5, no. 262, p. 1–7.

PAUL D. B., 1994, « Toward a Realistic Assessment of PKU Screening », *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1994, p. 322–328.

—, 2000, « PKU and Procreative Liberty: Historical and Ethical Considerations », *Ethical Issues in Health Care on the Frontiers of the Twenty-First Century*, S. Wear *et al. éds., Dordrecht, Kluwer Academic Publishers, p. 171–190.*

—, 2013, The PKU paradox: a short history of a genetic disease, Baltimore, Johns Hopkins University Press (Johns Hopkins biographies of disease).

PAUL N., 1998, « Incurable suffering from the “hiatus theoreticus”? Some epistemological problems in modern medicine and the clinical relevance of philosophy of medicine », *Theoretical Medicine and Bioethics*, vol. 19, no. 3, p. 229–251.

PAWSON T. et LINDING R., 2008, « Network medicine », *FEBS letters*, vol. 582, no. 8, p. 1266–1270.

PEARSON K., 1912, *Tuberculosis, Heredity and Environment*, London, Dulau.

PELLET J. *et al.,* 2010, « ViralORFeome: an integrated database to generate a versatile collection of viral ORFs »*, Nucleic Acids Research*, vol. 38, Database Issue, p. D371–D378.

PENROSE L. S., 1998, « Phenylketonuria–a problem in eugenics », *Annals of Human Genetics*, vol. 62, no. 3, p. 193–202.

PETRONIS A., 2010, « Epigenetics as a unifying principle in the aetiology of complex traits and diseases », *Nature*, vol. 465, no. 7299, p. 721–727.

PHILIPS P. C., 2008, « Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems », *Nature Reviews Genetics*, vol. 9, no. 11, p. 855–867.

PIRO R. M., 2012, « Network medicine: linking disorders », *Human Genetics*, no. 131, p. 1–10.

PLATT O. S. *et al.,* 1994, « Mortality In Sickle Cell Disease – Life Expectancy and Risk Factors for Early Death »*, New England Journal of Medicine*, vol. 330, no. 23, p. 1639–1644.

PORTELA A. et ESTELLER M., 2010, « Epigenetic modifications and human disease », *Nature Biotechnology*, vol. 28, no. 10, p. 1057–1068.

PORTIN P., 2002, « Historical Development of the Concept of the Gene », *Journal of Medicine and Philosophy*, vol. 27, no. 3, p. 257–286.

PRADEU T., 2007, *L'immunologie et la définition de l'identité biologique*, Université Panthéon-Sorbonne, Paris, France.

—, 2009, *Les limites du soi. Immunologie et identité biologique*, Paris, Presses Universitaires de Montréal, Montréal et Vrin.

—, 2010, « The organism in developmental systems theory », *Biological Theory*, vol. 5, no. 3, p. 216–222.

—, 2011, « Philosophie de la biologie », *Précis de philosophie des sciences*, A. Barberousse, D. Bonnay et M. Cozic éds., Paris, Vuibert, p. 378–403.

PRADEU T., JAEGER S. et VIVIER E., 2013, « The speed of change: towards a discontinuity theory of immunity? », *Nature Reviews Immunology*, vol. 13, no. 10, p. 764–769.

PRADEU Thomas, 2012, *The Limits of the Self: Immunology and Biological Identity*, Oxford University Press.

PRITCHARD J. K., 2002, « The allelic architecture of human disease genes: common disease-common variant... or not? », *Human Molecular Genetics*, vol. 11, no. 20, p. 2417–2423.

PUFFER R., 1944, *Familial Susceptibility to Tuberculosis: Its Importance as a Public Health Problem*, Cambridge, MA, Harvard University Press.

RAKYAN V. K. *et al.,* 2011, « Epigenome-wide association studies for common human diseases »*, Nature Reviews Genetics*, vol. 12, no. 8, p. 529–541.

RAM P. T., MENDELSOHN J. et MILLS G. B., 2012, « Bioinformatics and systems biology », *Molecular Oncology*, vol. 6, no. 2, p. 147–154.

RAMOZ N. *et al.,* 2002, « Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis »*, Nature Genetics*, vol. 32, no. 4, p. 579–581.

RAYMOND V., 1998, « Premier exemple de dominance autosomique spécifique de l'hétérozygote chez l'humain », *M/S : médecine/sciences*, vol. 14, no. 12, p. 1432–1434.

REDON R. *et al.,* 2006, « Global variation in copy number in the human genome »*, Nature*, vol. 444, no. 7118, p. 444–454.

REICH D. E. et LANDER E. S., 2001, « On the allelic spectrum of human disease », *Trends in Genetics*, vol. 17, no. 9, p. 502–510.

REZNEK L., 1987, *The nature of disease*, London ; New York, Routledge & Kegan Paul (Philosophical issues in science).

—, 1995, « Dis-Ease about Kinds: Reply to D'Amico », *Journal of Medicine and Philosophy*, vol. 20, no. 5, p. 571–584.

RICHARD H., à venir, *Diagnostic et dépistage prénatal: les maladies génétiques posent-elles un problème spécifique ?*, Thèse de doctorat, Université Panthéon-Sorbonne, Paris, France.

RICHARDS M., 2001, « How Distinctive is Genetic Information? », *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 32, no. 4, p. 663–687.

RISCH N. J. et MERIKANGAS K., 1996, « The Future of Genetic Studies of Complex Human Diseases », *Science*, vol. 273, no. 5281, p. 1516–1517.

ROUKOS D. H., 2011, « Networks medicine: from reductionism to evidence of complex dynamic biomolecular interactions », *Pharmacogenomics*, vol. 12, no. 5, p. 695–698.

ROUSSEAU F. et LAFLAMME N., 2003, « Génétique moléculaire humaine: des maladies monogéniques aux maladies complexes », *M/S: médecine sciences*, vol. 19, no. 10.

RUSE M., 1973, *The Philosophy of Biology*, London, Hutchinson University Library.

RUTTER M. et PICKLES A., 1991, « Person-environment interactions: concepts, mechanisms and implication for data analysis », *Conceptualization and measurement of organism-environment interaction*, R. Plomin et T. D. Wachs éds., Washington, D.C., American Psychological Association, p. 105–141.

SADEGH-ZADEH K., 2000, « Fuzzy Health, Illness, and Disease », *Journal of Medicine and Philosophy*, vol. 25, no. 5, p. 605–2013.

—, 2008, « The prototype resemblance theory of disease », *Journal of Medicine and Philosophy*, vol. 33, no. 2, p. 106–139.

—, 2011, *Handbook of the Analytic Philosophy of Medicine*, Dordrecht, Pays-Bas, Springer.

SANCHO-SHIMIZU V. *et al.,* 2011, « Herpes simplex encephalitis in children with autosomal recessive and dominant TRIF deficiency »*, The Journal of clinical investigation*, vol. 121, no. 12, p. 4889–4902.

SARKAR Sahotra, 1998, *Genetics and Reductionism*, Cambridge, Cambridge University Press.

SCHAFFNER K. F., 1980, « Theory Structure in the Biomedical Sciences », *Journal of Medicine and Philosophy*, vol. 5, no. 1, p. 57–97.

—, 1986, « Exemplar Reasoning About Biological Models and Diseases: A Relation Between the Philosophy of Medicine and Philosophy of Science », *Journal of Medicine and Philosophy*, vol. 11, no. 1, p. 63–80.

—, 1993a, *Discovery and Explanation in Biology and Medicine*, Chicago, The Chicago University Press.

—, 1993b, « Theory Structure, Reduction, and Disciplinary Integration in Biology », *Biology and Philosophy*, vol. 8, no. 3, p. 319–347.

—, 1998, « Genes, behavior and developmental emergentism: one process, indivisible? », *Philosophy of Science*, vol. 65, no. 2, p. 209–252.

—, 2006, « Reduction: The Cheshire Cat Problem and a Return to Roots », *Synthese*, vol. 151, no. 3, p. 377–402.

—, 2008, « Theories, Models, and Equations in Biology: The Heuristic Search for Emergent Simplifications in Neurobiology », *Philosophy of Science*, vol. 75, no. 5, p. 1008–1021.

—, 2011, « Reduction in Biology and Medicine », *Philosophy of Medicine*, Amsterdam ; Boston, Elsevier/North Holland (Handbook of the philosophy of science), v.16.

SCHAFFNER Kenneth F., 1993, *Discovery and Explanation in Biology and Medicine*, University of Chicago Press.

SCHWARTZ P. H., 2007, « Decision and discovery in defining “disease” »,, Springer.

SCRIVER C. R., 1995, « Whatever happened to PKU? », *Clinical biochemistry*, vol. 28, no. 2, p. 137–144.

—, 2007, « The PAH gene, phenylketonuria, and a paradigm shift », *Human mutation*, vol. 28, no. 9, p. 831–845.

SCRIVER C. R. et WATERS P. J., 1999, « Monogenic traits are not simple: lessons from phenylketonuria », *Trends in Genetics*, vol. 15, no. 7, p. 267–272.

SERJEANT G. R., 1981, « Observations on the epidemiology of sickle cell disease », *Transactions of the Royal Society of Tropical Medicine and Hygiene*, vol. 75, no. 2, p. 228–233.

SEVERINSEN M., 2001, « Principles Behind Definitions of Diseases – a Criticism of the Principle of Disease Mechanism and the Development of a Pragmatic Alternative », *Theoretical Medicine and Bioethics*, vol. 22, no. 4, p. 319–336.

SHARP D., 2005, « Genetic epidemiology: strengths, weaknesses, and opportunities », *The Lancet*, vol. 366, no. 9489, p. 880.

SHEA V., 2005, « Lumpers, splitters and Asperger syndrome », *Journal of Autism and Developmental Disorders*, vol. 35, no. 6, p. 871–872.

SHI X. *et al.,* 2013, « Long non-coding RNAs: a new frontier in the study of human diseases »*, Cancer letters*, vol. 339, no. 2, p. 159–166.

SILVERMAN E. K. et LOSCALZO J., 2012, « Network Medicine Approaches to the Genetics of Complex Diseases », *Discovery medicine*, vol. 14, no. 75, p. 143–152.

—, 2013, « Developing New Drug Treatments in the Era of Network Medicine », *Clinical pharmacology and therapeutics*, vol. 93, no. 1, p. 26–28.

SKINNER M. K., MANIKKAM M. et GUERRERO-BOSANGNA C., 2010, « Epigenetic transgenerational actions of environmental factors in disease etiology », *Trends in Endocrinology and Metabolism*, vol. 21, no. 4, p. 214–222.

SMART J. C., 1963, *Philosophy and Scientific Realism*, Routledge and Kegan Paul., London.

SMITH G. D. *et al.,* 2005, « Genetic epidemiology and public health: hope, hype, and future prospects »*, The Lancet*, vol. 366, no. 9495, p. 1484–1498.

SMITH K. C., 1992, « The new problem of genetics: A response to Gifford », *Biology and Philosophy*, vol. 7, no. 3, p. 331–348.

—, 2001, « A disease by any other name: musings on the concept of a genetic disease », *Medicine, Health Care and Philosophy*, vol. 4, no. 1, p. 19–30.

—, 2007, « Towards an Adequate Account of Genetic Disease », *Establishing Medical Reality - Essays in the Metaphysics and Epistemology of Biomedical Science*, H. Kincaid et J. McKitrick éds., Dordrecht, The Netherlands, Springer, p. 83–110.

SOBER E., 1988, « Apportioning Causal Responsibility », *The Journal of Philosophy*, vol. 85, no. 6, p. 303–318.

SOBER E. et WILSON D. S., 1994, « A Critical Review of Philosophical Work on the Units of Selection Problem », *Philosophy of Science*, vol. 61, no. 4, p. 534–555.

SPORNS O., 2012, « From simple graphs to the connectome: Networks in neuroimaging », *NeuroImage*, vol. 62, no. 2, p. 881–886.

STEGMANN U.E., 2012, « Varieties of parity », *Biology and Philosophy*, vol. 27, no. 6, p. 903–918.

STERELNY K., DICKINSON M. et SMITH K., 1996, « The extended replicator », *Biology and Philosophy*, vol. 11, no. 3, p. 377–403.

STEWART G. T., 1968, « Limitations of the germ theory », *The Lancet*, vol. 291, no. 7551, p. 1077–1081.

STOTZ K., GRIFFITHS P. E. et KNIGHT R.D., 2004, « How Biologists Conceptualize Genes: An Empirical Study », *Studies in History and Philosophy of Science Part C*, vol. 35, no. 4, p. 647–673.

STRASSER B. J., 2002, « Linus Pauling's "molecular diseases": Between history and memory », *American Journal of Medical Genetics*, vol. 115, no. 2, p. 83–93.

STROUSE J. J. *et al.,* 2008, « Hydroxyurea for Sickle Cell Disease: A Systematic Review for Efficacy and Toxicity in Children »*, Pediatrics*, vol. 122, no. 6, p. 1332–1342.

SULMASY D. P., 2005, « Diseases and natural kinds », *Theoretical medicine and bioethics*, vol. 26, no. 6, p. 487–513.

SUPPE F., 1972, « What's wrong with the received view on the structure of scientific theories? », *Philosophy of Science*, vol. 39, no. 1, p. 1–19.

—, 1977, « The search for philosophic understanding of scientific theories », *The Structure of Scientific Theories*, F. Suppe ed., Chicago, University of Illinois Press, p. 3–241.

SUPPES P., 1961, « A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences », *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, Netherlands, Springer, p. 163–177.

—, 1967, « What is a scientific theory? », *Philosophy of Science Today*, S. Morgenbesser ed., New York, Basic Books Inc., p. 55–67.

SUZUKI D. T. et KNUDTSON P., 1989, *Genethics: The Clash Between the New Genetics and Human Values*, Harvard, Harvard University Press.

SZASZ T. S., 1960, « The myth of mental illness », *American Psychologist*, vol. 15, no. 2, p. 113–118.

TABERY J., 2007, « Biometric and developmental gene–environment interactions: Looking back, moving forward », *Development and Psychopathology*, vol. 19, no. 04, p. 961–976.

—, 2009a, « Making sense of the nature-nurture debate - Review of Neven Sesardic (2005), Making sense of heritability », *Biology and Philosophy*, vol. 24, no. 5, p. 711–723.

—, 2009b, « From a Genetic Predisposition to an Interactive Predisposition: Rethinking the Ethical Implications of Screening for Gene-Environment Interactions », *Journal of Medicine and Philosophy*, vol. 34, no. 1, p. 27–48.

TABOR H. K., RISCH N. J. et MYERS R. M., 2002, « Candidate-gene approaches for studying complex genetic traits: practical considerations », *Nature Reviews Genetics*, vol. 3, no. 5, p. 391–397.

TENESA A. et HALEY C. S., 2013, « The heritability of human disease: estimation, uses and abuses », *Nature Reviews Genetics*, vol. 14, no. 2, p. 139–149.

TENNESSEN J. A. *et al.,* 2012, « Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes »*, Science*, vol. 337, no. 6090, p. 64–69.

THAGARD P., 1998, « Explaining disease: Correlations, causes, and mechanisms », *Minds and Machines*, vol. 8, no. 1, p. 61–78.

—, 2000, *How scientists explain disease*, Princeton, N.J., Princeton University Press.

—, 2006, « What is a medical theory? », *Multidisciplanary approaches to theory in medicine*, R. Paton et L. A. McNamara éds., Amsterdam, Elsevier, p. 47–62.

—, 2008, « Conceptual change in the history of science: Life, mind, and disease », *International handbook of research on conceptual change*, S. Vosniadou ed., London, Routledge, p. 374–387.

THE INTERNATIONAL HAPMAP CONSORTIUM, 2005, « A haplotype map of the human genome », *Nature*, vol. 437, no. 7063, p. 1299–1320.

THERY F., 2013, *L'importance biologique des ARN non codants : perspectives historique et philosophique.*, Thèse de doctorat, Université Panthéon-Sorbonne, Paris, France.

THOMAS D., 2010, « Gene-environment-wide association studies: emerging approaches », *Nature Reviews Genetics*, vol. 11, no. 4, p. 259–272.

THOMPSON R. P., 1989, *The Structure of Biological Theories*, New York, State University of New York Press.

—, 2010, « Causality, mathematical models and statistical association: dismantling evidence-based medicine », *Journal of Evaluation in Clinical Practice*, vol. 16, no. 2, p. 267–275.

—, 2011a, « Theories and models in medicine », *Philosophy of Medicine*, F. Gifford ed., Amsterdam ; Boston, Elsevier/North Holland (Hanbook of philosophy of science), v.16, p. 115–136.

—, 2011b, « Causality, theories and medicine », *Causality in the Sciences*, P. M. Illari, F. Russo et J. Williamson éds., Oxford University Press., Oxford, p. 25–44.

TIMP W. and FEINBERG A. P., 2013, « Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host », *Nature Reviews Cancer*, vol. 13, no. 7, p. 497–510.

TRABOULSI E., KOENEKOOP R. et STONE E.M., 2006, « Lumpers or splitters? The role of molecular diagnosis in Leber congenital amaurosis », *Ophtalmic genetics*, vol. 27, no. 4, p. 113–115.

TREVATHAN W., SMITH E. O. and MCKENNA J. J. éds., 2008, *Evolutionary medicine and health: new perspectives*, New York, Oxford University Press.

TSOU J. Y., 2013, « Depression and suicide are natural kinds: Implications for physician-assisted suicide », *International Journal of Law and Psychiatry*, vol. 36, no. 5–6, p. 461–470.

TURNBAUGH P. J. *et al.,* 2007, « The human microbiome project: exploring the microbial part of ourselves in a changing world »*, Nature*, vol. 449, no. 7164, p. 804–810.

TURNEY J. et TURNER J., 2000, « Predictive medicine, genetics and schizophrenia », *New Genetics and Society*, vol. 19, no. 1, p. 5–22.

UHER R., 2008, « Gene-environment interaction: overcoming methodological challenges », *Genetic effects on environmental vulnerability to disease*, M. Rutter ed., Chichester, UK, John Wiley and Sons Ltd, p. 13–30.

VALLE D., 2008, « Victor Almon McKusick, MD, 1921–2008, In Memoriam », *American Journal of Human Genetics*, vol. 83, no. 3, p. 301.

VANDAMME D. *et al.,* 2013, « Systems medicine: helping us understand the complexity of disease »*, QJM*, vol. 106, no. 10, p. 891–895.

VIRGIN H. W. et TODD J. A., 2011, « Metagenomics and personalized medicine », *Cell*, vol. 147, no. 1, p. 44–56.

VISSCHER P. M. *et al.,* 2012, « Five Years of GWAS Discovery », The American Journal of Human Genetics, vol. 90, no. 1, p. 7–24.

VISSCHER P. M., HILL W. G. et WRAY N. R., 2008, « Heritability in the genomics era — concepts and misconceptions », *Nature Reviews Genetics*, vol. 9, no. 4, p. 255–266.

DE VREESE L., à venir, « Rethinking the concept of disease debate: a pragmatist alternative ».

WAGNER A., 2003, « How the global structure of protein interaction networks evolves », *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, no. 1514, p. 457–466.

WAGNER A. et FELL D. A., 2001, « The small world inside large metabolic networks », *Proceedings of the Royal Society B: Biological Sciences*, vol. 268, no. 1478, p. 1803–1810.

WAKEFIELD J. C., 1992a, « The concept of mental disorder: On the boundary between biological facts and social values », *American Psychologist*, vol. 47, no. 3, p. 373–388.

—, 1992b, « Disorder as harmful dysfunction: A conceptual critique of DSM-III-R's definition of mental disorder », *Psychological Review*, vol. 99, no. 2, p. 232–247.

—, 2000, « Aristotle as sociobiologist: The "function of a human being" argument, black box essentialism, and the concept of mental disorder », *Philosophy, Psychiatry and Psychology*, vol. 7, , p. 17–44.

—, 2007, « The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis », *World Psychiatry*, vol. 6, no. 3, p. 149–156.

WALHOUT A. J. *et al.,* 2000, « Protein Interaction Mapping in C.elegans Using Proteins Involved in Vulval Development »*, Science*, vol. 287, no. 5450, p. 116–122.

WARD L. D. et KELLIS M., 2012, « Interpreting noncoding genetic variation in complex traits and human disease », *Nature biotechnology*, vol. 30, no. 11, p. 1095–1106.

WATERS C. K., 2007, « Causes that make a difference », *Journal of Philosophy*, vol. 104, no. 11, p. 551–579.

WATSON J. D., 1990, « The Human Genome Project: past, present, future », *Science*, vol. 248, no. 4951, p. 44–48.

WEBER M., 2005, *Philosophy of experimental biology*, Cambridge, Cambridge University Press.

WHITLEY R. J., 2006, « Herpes simplex encephalitis: adolescents and adults », *Antiviral research*, vol. 71, no. 2-3, p. 141–148.

WILLIAMS G. C. et NESSE R. M., 1991, « The dawn of Darwinian medicine », *The Quarterly review of biology*, vol. 66, no. 1, p. 1–22.

WINGERSON L., 1990, *Mapping our genes: the Genome Project and the future of medicine*, New York, Plume.

WOLKENHAUER O. *et al.,* 2013, « The road from systems biology to systems medicine », Pediatric research, vol. 73, 4 Pt 2, p. 502–507.

WOLKENHAUER O. et GREEN S., 2013, « The search for organizing principles as a cure against reductionism in systems medicine », *The FEBS journal*, vol. 280, no. 23, p. 5938–5948.

WOODWARD James, 2013, « Mechanistic Explanation: Its Scope and Limits », *Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 87, no. 1, p. 39–65.

WORBOYS M., 2000, *Spreading germs: disease theories and medical practice in Britain, 1865-1900*, Cambridge, Cambridge University Press.

—, 2007, « Was there a Bacteriological Revolution in late nineteenth-century medicine? », *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, vol. 38, no. 1, p. 20–42.

WORLD TECHNOLOGY EVALUATION CENTER, 2007, *Systems Biology: International Research and Development*, Dordrecht, The Netherlands, Springer.

WULFF H. R., 1984, « Comments on Hesslow's "What is a genetic disease?" », *Health, disease, and causal explanations in medicine*, L. Nordenfelt et B. I. B. Lindahl éds., Dordrecht, Pays-Bas, D. Reidel publ. Co, p. 195–197.

YADAV S. P., 2007, « The Wholeness in Suffix -omics, -omes, and the Word Om », *Journal of Biomolecular Techniques: JBT*, vol. 18, no. 5, p. 277.

YOOK S. H., JEONG H. et BARABÁSI A. L., 2002, « Modeling the Internet's large-scale topology », *Proceedings of the National Academy of Sciences*, vol. 99, no. 21, p. 13382–13386.

YOUNG S.E. *et al.,* 2006, « Interaction between MAO-A genotype and maltreatment in the risk for conduct disorder: Failure to confirm in adolescent patients », *American Journal of Psychiatry*, vol. 163, no. 6, p. 1019–1025.

YOXEN E., 1982, « Constructing genetic diseases », *The Problem of medical knowledge: examining the social construction of medicine*, P. Wright et A. Treacher éds., Edinburgh, Edinburgh University Press, p. 144–161.

ZACHAR P., 2000, « Psychiatric Disorders Are Not Natural Kinds », *Philosophy, Psychiatry, & Psychology*, vol. 7, no. 3, p. 167–182.

—, 2002, « The Practical Kinds Model as a Pragmatist Theory of Classification », *Philosophy, Psychiatry, & Psychology*, vol. 9, no. 3, p. 219–227.

ZACHAR P. et KENDLER K. S., 2007, « Psychiatric Disorders: A Conceptual Taxonomy », *American Journal of Psychiatry*, vol. 164, no. 4, p. 557–565.

ZAITLEN N. et KRAFT P., 2012, « Heritability in the genome-wide association era », *Human Genetics*, vol. 131, no. 10, p. 1655–1664.

ZAMPIERI F., 2009, « Origins and History of Darwinian Medicine », *Humana Mente*, no. 9, p. 13–38.

ZANZONI A., SOLER-LÓPEZ M. et ALOY P., 2009, « A network medicine approach to human disease », *FEBS letters*, vol. 583, no. 11, p. 1759–1765.

ZHANG S. Y. *et al.,* 2007, « TLR3 Deficiency in Patients with Herpes Simplex Encephalitis »*, Science*, vol. 317, no. 5844, p. 1522–1527.

ZHANG X. *et al.,* 2011, « The expanded human disease network combining protein–protein interaction information »*, European Journal of Human Genetics*, vol. 19, no. 7, p. 783–788.

ZHAO Y. et MOONEY S. D., 2012, « Functional organization and its implication in evolution of the human protein-protein interaction network », *BMC Genomics*, vol. 13, no. 1, p. 150.

ZUK O. *et al.,* 2012, « The mystery of missing heritability: Genetic interactions create phantom heritability »*, Proceedings of the National Academy of Sciences*, vol. 109, no. 4, p. 1193–1198.

Index des auteurs cités

A

Abecasis, G. R. 163
Abel, L. 278-279, 282-287, 289-294, 297-300, 303-304, 306-309, 311, 314-320, 323-327, 330-334, 336, 389, 433-435, 444
Abrahamsen, A. 325
Alberts, B. 18, 489, 491-492, 495, 500-501, 503, 505, 507
Alcaïs, A. 279, 293, 301-304, 308-310, 313, 322-323, 330
Al-Herz, W. 294
Al-Musen, S. 304
Altare, F. 292
Alter, A. 322, 323
Anderson, W. 287
Antman, E. 342, 401
Arrell, D. K. 342
Aschard H. 182
Auffray, C. 349, 417

B

Badano, L. 82, 92
Badger, J. H. 449
Baker A. 358
Barabási, A. L. 273, 339-342, 357, 359, 361, 363-367, 369-371, 375, 385, 387, 391-393, 418, 436
Bassaglia, Y. 18, 489
Beaudet, A. L. 187
Bechtel, W. 325, 415
Becker K. G. 380
Beckmann, J. S. 159
Berenson, A. E. 81
Bird, A. 184, 185
Blau, N. 90
Bodmer, W. F. 238
Boehnke, M. 150, 458
Bonabeau, E. 365, 369
Boorse, C. 469
Botkin, J. R. 126, 522
Botstein, D. 150, 457
Bowker, G. C. 41, 46, 47
Braillard, P. A. 343, 345, 359, 371
Bressler, J. 187
Bridges C. B. 159
Brigandt, I. 415
Broadbent, A. 234, 280, 482
Brown, W. M. 463, 464
Bruggeman, F. J. 346
Brunner, H. G. 406
Bruton, O. C. 291, 294
Burian, R.M. 41, 259
Burley, J. 125
Burton, P. R. 155
Bynum, W. F. 236

C

Campaner, R. 325
Caplan, A. L. 113
Capobianco, E. 349
Carter, K. C. 217, 280, 482
Carvunis, A.R. 343, 344
Casadevall, A. 283, 288, 318-319, 345
Casanova, J. L. 278, 279, 282-287, 289-295, 297-300, 303-304, 306-309, 311, 314-320, 323-327, 330-334, 336, 389, 433, 435, 444
Caspi, A. 178-180, 329
Casrouge, A. 296
Castro, O. 236
Cavalli-Sforza, L. 238
Chan, S. Y. 342, 360, 411

Chen, Z. 418
Childs, B. 352, 353, 357, 375
Choi, M. 163, 375
Chung, C. S. 256
Clayton, D. G. 152, 460
Clermont, G. 349, 350
Cohen, B. H. 288
Cohen, M. L. 288
Collins, F. S. 52, 125, 346
Cordell, H. J. 152, 169, 171, 172, 173, 460
Coventry, P. 41
Craddock, N. 160
Craver, C. 240, 325, 328, 487

D

Daly, M. 150, 458
Darden, L. 240, 325, 328
Dawkins, R. 201, 257
De Backer, P. 345
De Vreese, L. 473, 475
Dean, M. 301
Debret, G. 394-396
Dekeuwer, C. 42
Demenais, F. 292
Dempfle, A. 182, 183
Dessein, A. 292
Deutsch, J. 258
Dewan, A. 155, 256
Dickson, S. P. 162
Dietrich, M. R. 233
Dipple, K. M. 82, 91, 93, 112
Do, R. 163
Dobzhansky, T. 237
Duncan, C. J. 221
Dupré, J. 260, 321, 449

E

Ebrahim, S. 257
Edwards, J. H. 299
Eichler, E. E. 158
Engelhardt, H. T. 463, 469
Ereshefsky, M. 473, 475
Esteller, M. 447, 448, 451
Evans, A. S. 281

F

Falk, R. 40, 259
Falkow, S. 283, 290
Fang, F. S. 345
Feinberg, A. P. 187, 451, 452
Feingold, J. 150, 236, 457
Fell, D. A. 366
Ferguson-Smith, A. C. 451
Fieschi, C. 279, 304, 310, 323, 330
Fischer, R. A. 149, 152, 168, 500
Flores, M. 349, 350
Fox Keller, E. 40, 369, 370
Frigg, R. 212

G

Galas, D. J. 350
Gannett, L. 121-123
Garrod, A. 291, 355
Gaschignard, J. 322, 334
Gayon, J. 40, 41, 101-103, 143, 215, 259
Germain, P. L. 447
Gibson, G. 161, 164, 165
Giere, R. N. 212
Gifford, F. 97, 100, 111, 115, 219, 220, 221
Giroux, E. 468, 469, 470, 471
Glennan, S. 325
Godfrey-Smith, P. 193, 201
Goh, G. 163, 375
Goh, K. I. 357, 374, 377, 378, 380, 385, 388, 391
Gould, S. J. 145, 194, 239
Gray, R. D. 197-199, 202-204
Green, M. J. 126, 344
Green, S. 126, 344

Griffiths, P. E. 174, 193, 197-200, 201, 202, 203, 204, 259, 496
Grmek, M. D. 288
Guerrero-Bosangna, C. 187
Guo, Y. 296
Guttmacher, A. E. 346

H

Haberstick, B. C. 181
Hacein-Bey-Abina, S. 114, 324
Hacking, I. 477
Hagen, J. B. 348
Haley, C. S. 147
Hamosh, A. 55
Han, J. D. J. 370
Hanahan, D. 452
Harper, P. S. 40, 52
Harris, J. 125
Hartmann, S. 212
Hedgecoe, A. M. 41, 86-88
Hempel, C. G. 213
Herman, M. 296
Hesslow, G. 97, 98, 99
Hirschhorn, J. N. 150, 458
Hoedemaekers, R. 86
Hofmann, B. 470, 471
Holm, S. 126
Holtzman, N.A. 126
Hood, L. 125, 343, 348, 350, 418
Hooper, L. V. 449
Hudson, T. J. 347
Hull, D. L. 201
Hull, R. T. 97, 100, 104, 201
Huneman, P. 199, 412-414
Hunter, D. J. 178

I

Ilkilic, I. 126

J

Jablonka, E. 184
Jacunski, A. 401
Janjic, V. 373
Jensen, A. R. 145, 147, 175, 176, 205
Jeong, H. 364, 365
Jiang, H. 187, 451
Jimenez-Sanchez, G. 353, 357
Jordan, B. 141, 149, 158, 161, 163-164, 455,

K

Kakuk, P. 126, 127
Kathiresan, S. 163
Katsanis, N. 82, 92
Kavaloski, V. 215
Kellis, M. 448
Kendler, K. S. 485, 487
Kerr, A. 41, 86, 88
Kevles, D. J. 125
Kholodenko, B. 349
Khoury, M. 93, 256
Kim-Cohen, J. 181
Kincaid, H. 218, 247, 454, 464-467, 477
Kitano, H. 343, 367
Kitcher, P. 198, 203
Klein, C. 156
Klein, R. 155
Knight, R. D. 200, 201
Knudson, A. G. 466
Knudtson, P. 125
Kolch, W. 349
Kraft, P. 148, 156
Ku, C. S. 156
Kuhn, T. S. 213

L

Laflamme, N. 82, 93
Lakatos, I. 213
Lalouel, J. M. 299
Lamb, M. J. 184
Lander, E. S. 152, 500
Lange, M. 484

Laplane, L. 8, 215, 216
Lawrence, T. 295
Leavitt, J. W. 281
Lederberg, J. 288
Lee, P. 158, 347
Lee, S. G. 158, 347, 528
Lemoine, M. 7, 8, 214, 231, 232, 233, 234, 244, 461, 468, 470, 471, 473, 474, 475, 521, 528
Lesne, A. 344, 368, 394, 395, 408, 528
Levins, R. 192, 528
Levy, A. 415, 529
Lewens, T. 125, 529
Lewontin, R. C. 145, 147, 175, 176, 185, 186, 188, 190, 192, 194, 195, 197, 198, 199, 239, 522, 528, 529
Lindee, M. S. 39, 41, 59, 60, 86, 529, 532
Linding, E. 342, 534
Lippman, A. 85, 87, 529
Littman, D. R. 524
Liu, R. 301, 529
Lorne, M. C. 521
Loscalzo, J. 80, 342, 357, 360, 397, 398, 400, 401, 403, 408, 411, 512, 516, 529, 537
Lotteau, V. 417, 433, 532
Lottenberg, R. 236, 529
Lush, J. L. 143, 498, 529

M

Machamer, P. 9, 240, 325, 328, 529, 530
Mackie, J. L. 97, 98, 109, 530
Macpherson, A. J. 524
Magnus, D. 97, 101, 104, 113, 119, 120, 121, 122, 530
Manikkam, M. 187, 538
Maniolo, T. A. 155, 157, 158, 530
Maubaret, C. G. 69, 530
Mayr, E. 237, 530
McCabe, E. R. B. 82, 91, 93, 112, 518
McCarty, M. I. 156, 530
McKenna, J. J. 540
McKusick, V. A. 35, 52, 53, 54, 55, 56, 58, 62, 64, 65, 66, 71, 75, 76, 78, 80, 82, 83, 95, 503, 517, 523, 530, 541
Meaney, M. J. 184, 187, 530
Mears, J. G. 236
Merikangas, K. 152
Merlin, F. 203, 204
Mill, J. S. 97
Miller, F. A. 80
Miller, L. H. 301
Mira M. 322, 511
Mitchell, S. D. 233, 436
Montoya, J. M. 364, 412
Mooney, S. D. 366
Moore, J. H. 53, 171, 173
Morange, M. 7, 184
Morton, N. E. 256
Moulin, A. M. 284
Mueller, R. 41
Munson, R. 462, 463
Murphy, D. 325, 471
Myers, R. M. 152

N

Nacher, J. C. 401
Nagel, E. 104, 213
Navratil, V. 418, 420, 421, 424, 425, 427-429, 431-435
Neel, J. V. 256
Nelkin, D. 86
Nelson, M. R. 163
Nesse, R. M. 237, 239
Neumann-Held, E. M. 259
Nicholson, J. K. 449
Nordenfelt, L. 469

O

Ohno, S. 447
Oltvai, Z. N. 340, 361, 363, 367

Onnela, J. P. 364
Oti, M. 406
Oyama, S. 192, 197, 204

P

Park, J. 405
Paul, D. B. . 41
Paul, N. 454
Pawson, T. 342
Pearson, K. 291
Pellet, J. 418
Penrose, L. S. 59
Petronis, A. 187, 451
Philips, P. C. 168
Pickles, A. 177
Pickstone, J. V. 41
Piro, R. M. 342
Pirofski, L. A. 283, 288, 319
Platt, O. S. 236
Portela, A. 451
Portin, P. 258
Pradeu, T. 192-197, 257, 285, 290
Pritchard, J. K. 161
Puffer, R. 291

Q

Quintana-Murci, L. 279, 308, 309, 313

R

Rabourdin-Combe, C. 417, 433
Rakyan, V. K. 182
Ram, P. 348
Ramoz, N. 294
Raymond, V. 70
Redon, R. 159
Reich, D. E. 152, 500
Reznek, L. 475- 478, 480, 481, 484
Richard, H. 126
Richards, M. 126
Risch, N. J. 150, 152, 457
Roukos, D. H. 342
Rousseau, F. 82, 93
Ruse, M. 213
Rutter, M. 177

S

Sadegh-Zadeh, K. 28, 214, 219, 220, 223, 247
Sancho-Shimizu, V. 297
Sarkar, S. 102
Schaffner, K. F. 28, 198, 210, 211, 214, 219, 220, 247
Schull, W. J. 256
Schwartz, J. M. 401
Schwartz, P. H. 473
Scott, S. 221
Scriver, C. R. 41, 82, 89, 90
Serjeant, G. R. 236
Severinsen, M. 482
Sharp, D. 257
Shea, V. 81
Shi, X. 448
Silverman, E. K. 342, 401
Skinner, M. K. 187
Smart, J. C. 213
Smith, E. O.17, 97, 104, 105, 106, 111, 112, 115, 116, 117, 118, 119, 140, 148, 257, 287, 290
Smith, G. D. 257
Smith, K. C. 17, 97, 104-106, 111-112, 115-119, 140
Sober, E. 99, 257
Sporns, O. 412
Star, S. L. 41, 46, 47
Stearns, S. C. 237
Stegmann, U. E. 198
Sterelny, K. 203
Stewart, G. T. 281
Stotz, K. 259, 496
Strasser, B. J. 41
Strouse, J. L. 236
Sulmasy, D. P. 477

Suppe, F. 214
Suppes, P. 212, 214
Suzuki, D. 125
Szasz, T. S. 470, 475

T

Tabery, J. 174, 175, 176, 177, 184, 188, 329
Tabor, H. K. 152, 460
Tenesa, A. 147
Tennessen, J. A. 163
Terzic, A. 342
Thagard, P. 28, 210, 214, 223-230, 247, 263-264
The International HapMap Consortium 154, 159
Théry, F. 447
Thomas, D. 182
Thompson, R. P. 28, 210, 211, 214, 219-222, 247
Timp,W. 451, 452
Todd, J. A. 450
Traboulsi, E. 81
Trevathan, W. 237
Tsou, J. Y. 479
Turnbaugh, P. J. 278
Turner, J. 88
Turney, J. 88

U

Uher, R. 183

V

Valle, D. 52, 352, 353, 357, 375
Van Fraassen, B. C. 121, 232
Vandamme, D. 349
Virgin, H. W. 450
Visscher, P. M. 148, 155, 158

W

Wagner, A. 366
Wakefield, J. C. 469, 470
Walhout, A. J. 375
Ward, L. 448
Waters, C. K. 41, 82, 89, 90, 115
Waters, P. J. 41, 82, 89, 90, 115
Watson, J. D. 352
Weber, M. 199, 200
Weinberg, R. A. 452
Westerhoff, H. V. 346
Whitley, R. 296
Williams, G. C. 171, 237, 239
Williams, S. M. 171, 237, 239
Wilson, D. S. 257, 487
Wingerson, L. 352
Wolkenhauer, O. 344, 349-351
Woodward, J. 414, 415
Worboys, M. 280
Wray, N. R. 158
Wulff, W. R. 97, 108-111, 117

Y

Yadav, S. P. 346
Yook, S. H. 364
Young, S. E. 181
Yoxen, E. 39, 59

Z

Zachar, P. 485, 486, 487
Zaitlen, N. 148, 156
Zampieri, F. 236
Zanzoni, A. 342
Zhang, S. Y. 296
Zhang, X. 380
Zhao, Y. 366
Zuk, O. 167, 168, 172

Table des matières

Remerciements **7**

Sommaire **11**

Liste des figures **12**

Liste des tableaux **17**

Avertissements **18**

Liste des abréviations **19**

Introduction **21**

Le paradoxe de la génétique médicale contemporaine **21**

Enjeux du paradoxe de la génétique médicale **23**

Une résolution traditionnelle insatisfaisante du paradoxe de la génétique médicale **25**

L'hypothèse d'une théorie génétique de la maladie **26**

Identifier les conditions de possibilité d'une théorie génétique de la maladie **27**

Tester notre alternative **29**

Thèses soutenues et plan de l'argumentation **31**

Partie 1 : Du concept de maladie génétique à la généticisation des maladies **35**

Chapitre 1 : Problèmes scientifiques soulevés par le concept de maladie génétique **39**

1.1. Le concept de maladie génétique au prisme des classifications 39

1.1.1. Etudier l'histoire d'un concept à travers l'histoire de sa classification 39

1.1.2. Quelle place pour les maladies génétiques au sein de la Classification Internationale des Maladies (CIM-10)? 43

1.1.2.1. Une brève histoire de la Classification Internationale des Maladies 43

1.1.2.2. Présentation de la CIM -10 44

1.1.2.3. Une classification statistique et pragmatique, donc incomplète et ambiguë 45
1.1.2.4. Les maladies génétiques dans la CIM-10, une dispersion étonnante 48
1.1.2.5. Les raisons de la dispersion 50

1.2. Les maladies génétiques dans la base de données *Online Mendelian Inheritance on Man* (OMIM) 52

1.2.1. OMIM, classification « officieuse » des maladies génétiques 52
1.2.1.1. Victor McKusick, fondateur d'OMIM 52
1.2.1.2. La naissance de MIM 52
1.2.1.3. De MIM à OMIM, du catalogue à la base de données 54
1.2.2. Description d'une entrée dans OMIM 56
1.2.3. Le modèle de la maladie génétique aux débuts d'OMIM 58

1.3. Les principales transformations d'OMIM – bouleversement de l'hérédité mendélienne et extension du système de classification aux maladies non mendéliennes 60

1.3.1. Le système de classement des phénotypes : un classement en fonction du mode de transmission de la maladie 60
1.3.2. La découverte et l'inclusion des modes d'hérédité non mendélienne 62
1.3.2.1. L'inclusion des phénotypes liés à l'Y, une continuité logique 62
1.3.2.2. L'inclusion des phénotypes mitochondriaux, un problème plus délicat 63
1.3.2.3. Les anomalies chromosomiques : présentes mais invisibles 64
1.3.3. La remise en question des concepts de dominance et de récessivité dans l'hérédité mendélienne 65
1.3.3.1. Problème du niveau d'analyse du phénotype et de la semi-dominance 66
1.3.3.2. Problème de l'intervention des facteurs environnementaux 67
1.3.3.3. Problème de la pénétrance incomplète et de l'expressivité variable 68
1.3.3.4. Problème de la dominance autosomique spécifique de l'hétérozygote 70
1.3.4. L'ouverture d'OMIM aux maladies polygéniques 70

1.4. Les principales transformations d'OMIM – d'un classement des maladies à un classement des variants alléliques 71

1.4.1. Le système des symboles : un classement en fonction de la nature de l'objet référencé 72
1.4.2. Le concept de variant allélique 74
1.4.3. Hétérogénéité génétique, hétérogénéité allélique et pléiotropie 76
1.4.3.1. Le phénomène « un gène-plusieurs phénotypes » (pléiotropie ou hétérogénéité phénotypique) 76

1.4.3.2. Le phénomène « un phénotype, plusieurs gènes » (hétérogénéité génétique) 77
1.4.3.3. Le phénomène « plusieurs allèles, une même maladie » (hétérogénéité allélique) 77
1.4.4. Deux conséquences de l'ouverture d'OMIM aux maladies polyfactorielles 78
1.4.4.1. Vers une classification moléculaire des maladies génétiques ? 79
1.4.4.2. Combien de maladies génétiques ? Combien de gènes de la maladie ? 81
1.5. Le concept de maladie génétique aujourd'hui – entre extension et dissolution ..82
1.5.1. Les maladies génétiques dans OMIM – bilan et perspectives 82
1.5.2. La généticisation des maladies 85
1.5.3. La dissolution du concept de maladie génétique 88
1.5.4. La notion de continuum génétique, révélateur des ambiguïtés de la génétique médicale 92
1.6. Conclusion du chapitre : le paradoxe de la génétique médicale contemporaine .95
Chapitre 2 : Problèmes philosophiques soulevés par le concept de maladie génétique : généticisation, génocentrisme et problème de la sélection causale 97
2.1. Définir la maladie génétique, une instance du problème de la sélection causale 97
2.1.1. Le problème de la sélection causale 97
2.1.2. Une stratégie commune : le rejet de la généticisation comme génocentrisme 101
2.1.3. Une stratégie commune : restreindre le concept de maladie génétique 105
2.2. Trois approches pour résoudre le problème de la sélection causale 107
2.2.1. L'approche cause/condition : définir la maladie génétique en soi 108
2.2.2. L'approche différentialiste : définir la maladie génétique par rapport à une population de contraste 115
2.2.3. L'approche pragmatique : définir la maladie génétique par rapport à un contexte explicatif 121
2.3. Il faut abandonner le concept de maladie génétique 123
2.3.1. Il n'y a pas de critère pour définir objectivement le concept de maladie génétique 123
2.3.2. Des conséquences scientifiques et sociales limitées 124

2.4. Il faut abandonner le problème de la sélection causale 128

2.4.1. Résoudre le problème de la sélection causale n'est d'aucune utilité en médecine 128

2.4.2. Résoudre le problème de la sélection causale ne fournit pas une explication de la maladie 129

2.4.3. Le problème de la sélection causale simplifie le contexte causal 130

2.4.4. Le problème de la sélection causale devrait être abandonné pour interpréter de façon différente et pertinente la généticisation 131

2.5. Conclusion du chapitre : changer l'approche de la généticisation 132

Partie 2 : De la généticisation des maladies aux conditions de possibilité et aux critères d'une théorie génétique de la maladie 133

Chapitre 3 : Comprendre la généticisation au sein d'une explication interactionniste de la maladie 139

3.1. Du concept d'héritabilité au cadre interactionniste de la génétique quantitative 142

3.1.1. Origine et définition du concept d'héritabilité 143

3.1.2. L'interprétation du concept d'héritabilité 146

3.1.2.1. L'héritabilité est un paramètre statistique relatif à une population et à un environnement donnés 146

3.1.2.2. Une héritabilité élevée n'implique pas la détermination génétique du trait 147

3.1.2.3. La définition de l'héritabilité ne doit pas être confondue avec les méthodes d'estimation de l'héritabilité 147

3.2. Le problème de l'héritabilité perdue dans l'explication des maladies 149

3.2.1. Méthodes d'identification des gènes de la maladie – des succès des maladies monogéniques au « cauchemar » des maladies complexes 149

3.2.2. Les associations pangénomiques, l'hypothèse « variant commun-maladie commune » 152

3.2.3. À la recherche de « l'héritabilité perdue » 158

3.2.3.1. Une autre source de variation commune 159

3.2.3.2. D'autres hypothèses sur l'architecture allélique de la maladie 160

3.2.3.3. Le modèle « de l'héritabilité au sens large » 164

3.3. Deux concepts d'interaction 166

3.3.1. Les interactions gène x gène ou l'épistasie 166

3.3.2. Les interactions gène x environnement 174

3.3.2.1. Deux concepts d'interaction gène x environnement 174

3.3.2.2. Les difficultés inhérentes à la détection des interactions gène x environnement et les problèmes posés par les associations pangénomiques 178

3.3.2.3. Les interactions gène x environnement remettent-elles en cause le concept d'héritabilité ? 182

3.3.3. Que nous apprend le cadre interactionniste de la génétique quantitative ? 188

3.4. L'interactionnisme co-constructionniste 192

3.4.1. Contexte et objet de l'interactionnisme co-constructionniste 193

3.4.1.1. Internalisme développemental et externalisme évolutif 193

3.4.1.2. L'organisme construit son environnement et l'environnement construit l'organisme 195

3.4.2. La thèse de la parité causale 197

3.4.3. La thèse de la parité causale est-elle applicable ? 202

3.4.4. Que nous apprennent l'interactionnisme co-constructionniste et la thèse de la parité causale ? 204

3.5. Conclusion du chapitre : l'hypothèse d'une théorie génétique de la maladie 205

Chapitre 4 : Établir des critères opérationnels pour identifier une théorie génétique 209

4.1. Pourquoi faire l'hypothèse d'une « théorie » génétique de la maladie ? 212

4.1.1. Rejet de la conception syntaxique des théories scientifiques 213

4.1.2. Examen des critères communément évoqués pour définir une théorie scientifique 214

4.1.3. Une conception déflationniste des théories scientifiques – critères de Laplane 215

4.1.4. Avantages d'une conception déflationniste des théories scientifiques 217

4.2. Qu'est-ce qu'une théorie médicale explique ? 219

4.2.1. L'objet des théories médicales est l'explication de la maladie 219

4.2.2. Différents niveaux de théorie médicale : théorie de la maladie, d'une classe de maladie et d'une maladie 223

4.2.3. Quatre cibles explicatives communes : causes, symptômes, évolution, traitement 226

4.3. Comment une théorie médicale explique-t-elle ? 231

4.3.1. Le pluralisme explicatif en médecine 231
4.3.2. L'explication épidémiologique 234
4.3.3. L'explication évolutionnaire 236
4.3.4. L'explication physiopathologique 240
4.3.5. L'explication clinique, lieu d'instanciation des différents types d'explications . 242
4.3.6. Il n'y a pas de hiérarchie entre les différents types d'explication en médecine 244

4.4. Jusqu'à quel point une théorie médicale doit-elle expliquer ? 245

4.4.1. Définition d'une théorie médicale a minima et avantages de notre conception 245
4.4.2. Trois classes de faits supplémentaires pour évaluer la précision d'une théorie médicale 248
4.4.2.1. La variabilité interindividuelle 249
4.4.2.2. La variabilité intra-individuelle 250
4.4.2.3. La comorbidité 251
4.4.2.4. Trois arguments en faveur de ces trois classes de faits 251
4.4.3. La notion de théorie partielle 253

4.5. A quoi ressemblerait une théorie génétique ? 254

4.5.1. Les gènes comme explanans dans les explications évolutionnaire, épidémiologique et physiopathologique d'une maladie 255
4.5.2. Le gène, concept en tension, qui permet le pluralisme explicatif 258
4.5.3. Critères et typologies des théories génétiques possibles 260

4.6. Conclusion du chapitre : la généticisation révèle-t-elle une théorie génétique des maladies ou une théorie génétique de la maladie ? 265

Partie 3 : Deux exemples contemporains de théorie génétique 269

Chapitre 5 : La théorie génétique des maladies infectieuses, un exemple de théorie génétique régionale 277

5.1. Pourquoi faire l'hypothèse d'une théorie génétique des maladies infectieuses ? 279

5.1.1. La variabilité interindividuelle aux infections, un manque explicatif de la théorie des germes 280

5.1.2. La place des théories microbienne, immunologique et écologique dans l'explication de la variabilité interindividuelle aux infections 282

5.1.2.1. La « théorie microbienne » n'explique pas la variabilité interindividuelle aux infections lorsque deux individus sont exposés à une même souche pathogène 282

5.1.2.2. La « théorie immunologique » n'explique pas les échecs de la vaccination ni les primo-infections infantiles potentiellement mortelles 284

5.1.2.3. La « théorie écologique » ne permet pas de comprendre la variabilité interindividuelle aux infections lorsque deux individus sont exposés au même pathogène dans un même environnement 286

5.1.3. De l'hypothèse génétique à la théorie génétique des maladies infectieuses ... 291

5.2. Présentation de la théorie génétique des maladies infectieuses 293

5.2.1. Quatre mécanismes génétiques communs non mutuellement exclusifs 293

5.2.1.1. La prédisposition mendélienne à plusieurs infections et les immunodéficiences primaires conventionnelles 294

5.2.1.2. La prédisposition mendélienne à une infection ou les nouvelles immunodéficiences primaires 295

5.2.1.3. Le concept de gène majeur / de résistance à une infection 299

5.2.1.4. La prédisposition polygénique à une ou plusieurs infections 301

5.2.2. Des mécanismes génétiques communs au continuum génétique de la maladie infectieuse 303

5.2.3. L'âge de début de la maladie, une variable prédictive de l'architecture génétique des maladies infectieuses 307

5.3. En quoi la théorie génétique des maladies infectieuses est-elle un exemple de théorie génétique régionale ? 314

5.3.1. La TGMI est-elle une théorie partielle qui s'intègre dans une explication interactionniste co-constructionniste des maladies infectieuses ? 314

5.3.2. La TGMI est-elle une théorie régionale qui explique les quatre cibles explicatives des maladies infectieuses ? 321
5.3.3. Quel type d'explication mobilise la TGMI ? 325
5.3.3.1. Rappels sur l'explication mécaniste 325
5.3.3.2. Première objection – des mécanismes insuffisamment détaillés ? 326
5.3.3.3. Deuxième objection – Des mécanismes fondés sur des concepts de l'interactionnisme quantitatif 328
5.3.3.4. Une explication mécaniste cohérente avec des données évolutionnaires et épidémiologiques 331
5.3.4. En quoi la TGMI explique-t-elle la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité des maladies infectieuses ? 331

5.4. Conclusion du chapitre : de la théorie génétique des maladies infectieuses à une théorie génétique des maladies ? 335

Chapitre 6 : La médecine des réseaux, un cadre pour une théorie génétique de la maladie ? 339

6.1. Les origines de la médecine des réseaux 340

6.1.1. Un nouveau contexte : la biologie et la médecine des systèmes 343
6.1.2. Un nouveau concept : les « gènes humains de la maladie » 351
6.1.3. Une nouvelle boîte à outils : la théorie des réseaux 357
6.1.4. La médecine des réseaux, « une approche de la maladie humaine fondée sur les réseaux » 370

6.2. Construire et analyser le diseasome 373

6.2.1. Construire le diseasome 375
6.2.2. Analyser globalement le diseasome comme un réseau invariant d'échelle 380
6.2.3. Analyser localement le diseasome : l'hypothèse locale 382
6.2.4. Analyser le diseasome au sein de l'interactome : propriétés des gènes humains de la maladie 384

6.3. En quoi la médecine des réseaux est-elle un cadre pour une théorie génétique générale de la maladie ? 388

6.3.1. La médecine des réseaux, une théorie générale interactionniste et co-constructionniste qui explique les causes, les symptômes, l'évolution et le traitement de la maladie 389
6.3.1.1. Chaque maladie peut être définie par son module 389

6.3.1.2. Mais chaque individu a un module de la maladie spécifique 392
6.3.1.3. La médecine des réseaux, un cadre interactionniste co-constructionniste 396
6.3.1.4. La médecine des réseaux est-elle une théorie générale qui explique les quatre cibles de la maladie ? 399
6.3.2. La théorie génétique de la maladie, une théorie partielle au sein de la médecine des réseaux 402
6.3.2.1. Expliquer l'origine génétique commune des maladies 402
6.3.2.2. Redéfinir la distinction entre maladies monogéniques et maladies polygéniques 406
6.3.2.3. Prédire de nouveaux gènes de la maladie 409
6.3.3. Quel type d'explication mobilise la médecine des réseaux ? 410
6.3.3.1. Explication topologique 411
6.3.3.2. Explication mécaniste 413
6.3.4. En quoi la théorie génétique de la médecine des réseaux explique-t-elle la variabilité interindividuelle, la variabilité intra-individuelle et la comorbidité ? 415

6.4. Les maladies infectieuses du point de vue de la médecine des réseaux 416

6.4.1. L'approche systémique des infections virales 416
6.4.2. Quand l'infectome rencontre le diseasome 420
6.4.2.1. L'analyse du réseau de l'infectome humain 423
6.4.2.2. L'analyse du réseau de l'infectome-diseasome humain 426
6.4.2.3. Conclusions de l'analyse du réseau de l'infectome humain et du réseau de l'infectome-diseasome humain 429
6.4.3. Les maladies infectieuses au sein de la TGMI et au sein de la médecine des réseaux 430

6.5. Conclusion du chapitre : limites méthodologiques et directions futures de la théorie génétique de la médecine des réseaux 433

Conclusion 437

1. La coexistence d'une pluralité de théories génétiques est nécessaire pour résoudre le paradoxe de la génétique médicale contemporaine 437

1.1. L'approche du paradoxe par la sélection causale est une impasse 437

1.2. Une multiplicité de théories génétiques possibles 439

1.3. Une multiplicité de théories génétiques de fait 442

1.4. La nécessaire coexistence de théories génétiques variées 443

2. Trois autres sources de complexité pour la génétique médicale 445

2.1. Les fonctions du génome non codant 445

2.2. Le rôle du microbiome 447

2.3. La place de l'épigénétique 448

3. Le concept de théorie médicale et la philosophie de la médecine 451

Annexes 453

Annexe 1 : (Chapitre 3) – Méthodes d'identification des variants alléliques – des maladies monogéniques au cauchemar des maladies complexes 453

Annexe 2 : (Chapitre 4) Prémisses nécessaires pour accepter le concept de théorie médicale 459

1. Le concept de théorie médicale suppose de reconnaître que la constitution du savoir médical n'est pas entièrement subordonnée à la pratique clinique 460

2. Le concept de théorie médicale suppose qu'on puisse distinguer – au moins provisoirement – les faits des valeurs dans une analyse des concepts de santé et de maladie 466

3. Le concept de théorie médicale ne présuppose pas qu'on définisse les maladies comme des espèces naturelles 473

Glossaire 487

Bibliographie 507

Index des auteurs cités 541

Y a-t-il une théorie génétique de la maladie ?

Résumé : Alors qu'il n'existe pas de définition consensuelle du concept de maladie génétique, ce concept s'est progressivement élargi pour désigner des maladies communes, non héréditaires, non mendéliennes et polygéniques, aboutissant à une généticisation des maladies. Pour résoudre ce paradoxe de la génétique médicale contemporaine, les philosophes réfutent généralement cette généticisation comme une extension génocentriste abusive du concept de maladie génétique et cherchent à redéfinir un concept plus strict de maladie génétique. Nous montrons que cette stratégie échoue et proposons au contraire d'abandonner le concept de maladie génétique et de supposer que la généticisation révèle l'élaboration d'une explication du rôle commun des gènes dans toutes les maladies, que nous appelons une « théorie génétique de la maladie ». Nous définissons les conditions de possibilité et les critères nécessaires d'une théorie génétique *a minima* et aboutissons à un spectre des théories génétiques possibles. Nous proposons alors de tester si la généticisation des maladies révèle plutôt une *théorie génétique des maladies*, c'est à dire un ensemble de théories génétiques spécifiques à chaque classe de maladie, ou *une théorie génétique de la maladie*, reposant sur une définition générale de la maladie qui unifie le rôle commun des gènes dans toutes les maladies. Pour ce faire, nous analysons deux exemples de théories génétiques contemporaines : la théorie génétique des maladies infectieuses et la théorie génétique de la médecine des réseaux. Nous concluons à la coexistence nécessaire de plusieurs formes de théories génétiques dans la littérature biomédicale contemporaine.

Mots-clés : Maladie génétique, Généticisation, Sélection causale, Théorie médicale, Mécanisme, Explication, Maladies Infectieuses, Médecine des systèmes, Médecine des réseaux, Philosophie de la Médecine, Philosophie des Sciences

Is there a genetic theory of disease ?

Summary: While there is no consensual definition of the concept of genetic disease, this concept has gradually extended to designate common, non-hereditary, non-Mendelian, polygenic diseases, leading to the geneticization of diseases. In order to solve this paradox of the contemporary medical genetics, philosophers usually discard geneticization as an inappropriate genocentrist extension of the concept of genetic disease and attempt to define a stricter concept of genetic disease. We demonstrate the failure of this strategy and argue on the contrary that we should give up the concept of genetic disease and understand geneticization as the elaboration of an explanation of the common role of genes in diseases, what we call "a genetic theory of disease". We define the conditions of possibility and the necessary criteria for a genetic theory *a minima* and end up with describing the spectrum of potential genetic theories. We then suggest to test whether geneticization of diseases reveals rather *a genetic theory of diseases*, that is, a set of genetic theories specific to each class of disease, or *a genetic theory of disease*, that is, a general definition of disease unifying the common role of genes in disease explanations. In order to do so, we analyze two examples of contemporary genetic theories: the genetic theory of infectious diseases and the genetic theory of network medicine. We conclude that several forms of genetic theories coexist in the contemporary biomedical literature and that this coexistence is necessary.

Key words: Genetic disease, Geneticization, Causal Selection, Medical Theory, Mechanism, Explanation, Infectious Diseases, Systems medicine, Network medicine, Philosophy of medicine, Philosophy of science

Discipline : Philosophie
Ecole Doctorale : École Doctorale de Philosophie, Université Paris 1 Panthéon-Sorbonne, 1 rue d'Ulm, 75005 Paris.
Equipe d'accueil : Institut d'Histoire et de Philosophie des Sciences et des techniques (UMR8590), 13 rue du Four, 75006 Paris.